

**PATENT APPLICATION**  
**NUCLEIC ACIDS THAT CONTROL PLANT DEVELOPMENT**

Inventor(s):

Robert L. Fischer, a citizen of the United States  
residing at: 1423 Scott Street, El Cerrito, CA 94530

Yeonhee Choi, a citizen of South Korea  
residing at: 4 Anchor Dr. #442, Emeryville, CA 94608

Mike Hannon, a citizen of the United States  
residing at: 309 Pearl Drive, Livermore, CA 94550

Jack Kishiro Okamuro, a citizen of the United States  
residing at: 307 Ocho Rios Way, Oak Park, CA 91377

Tatiana Valerievna Tatarinova, a citizen of Russia  
residing at 836 North Beaudry Avenue Apt. 2, Los Angeles, CA 90012

Assignee:

THE REGENTS OF THE UNIVERSITY OF CALIFORNIA  
1111 Franklin Street, Fifth Floor  
Oakland, CA 94607-5200

Entity:

TOWNSEND and TOWNSEND and CREW LLP  
Two Embarcadero Center, 8<sup>th</sup> Floor  
San Francisco, California 94111-3834  
Tel: 415-576-0200

09840743-042301

# NUCLEIC ACIDS THAT CONTROL PLANT DEVELOPMENT

## CROSS-REFERENCES TO RELATED APPLICATIONS

5 This application is a continuation-in-part of U.S. Patent Application No. 09/553,690, filed April 21, 2000, the contents of which are incorporated by reference.

## STATEMENT AS TO RIGHTS TO INVENTIONS MADE UNDER FEDERALLY SPONSORED RESEARCH AND DEVELOPMENT

10 This invention was made with Government support under Grant No. 97-35304-4941, awarded by the United States Department of Agriculture. The government has certain rights in this invention.

## FIELD OF THE INVENTION

15 This invention is directed to plant genetic engineering. It relates to, for example, modulating seed (and in particular endosperm, embryo and seed coat) development, flowering time, chromosomal DNA methylation and modulating transcription in plants.

## BACKGROUND OF THE INVENTION

20 A fundamental problem in biology is to understand how seed development. In flowering plants, the ovule generates the female gametophyte, which is composed of egg, central, synergid and antipodal cells (Reiser, *et al.*, *Plant Cell*, 1291-1301 (1993)). All are haploid except the central cell which contains two daughter nuclei that fuse prior to fertilization. One sperm nucleus fertilizes the egg to form the zygote, 25 whereas another sperm nucleus fuses with the diploid central cell nucleus to form the triploid endosperm nucleus (van Went, *et al.*, *Embryology of Angiosperms*, pp. 273-318 (1984)). The two fertilization products undergo distinct patterns of development. In Arabidopsis, the embryo passes through a series of stages that have been defined morphologically as preglobular, globular, heart, cotyledon and maturation (Goldberg, R. 30 B., *et al.*, *Science* (1994) 266: 605-614; Mansfield, S. G., *et al.*, *Arabidopsis: An Atlas of Morphology and Development*, pp. 367-383 (1994)). The primary endosperm nucleus undergoes a series of mitotic divisions to produce nuclei that migrate into the expanding central cell (Mansfield, S. G., *et al.*, *Arab Inf Serv* 27: 53-64 (1990); Webb, M. C., *et al.*, *Planta* 184:187-195 (1991)). Cytokinesis sequesters endosperm cytoplasm and nuclei

into discrete cells (Mansfield, S. G., *et al.*, *Arab Inf Serv* 27:65-72 (1990)) that produce storage proteins, starch, and lipids which support embryo growth (Lopes, M. A. *et al.*, *Plant Cell* 5:1383-1399 (1993)). Fertilization also activates development of the integument cell layers of the ovule that become the seed coat, and induces the ovary to grow and form the fruit, or silique, in *Arabidopsis*.

Of particular interest are recent discoveries of genes that control seed, and in particular endosperm, development. For instance, *MEDEA* (*MEA*) (also known as *FIE1* (see, e.g., copending U.S. patent application 09/071,838) and F644 (see, e.g., Kiyosue T, *et al.* (1999) *Proc Natl Acad Sci U S A* 96(7):4186-91) encodes an *Arabidopsis* SET domain polycomb protein that appears to play a role in endosperm development. Inheritance of a maternal loss-of-function *mea* allele results in embryo abortion and prolonged endosperm production, irrespective of the genotype of the paternal allele. Thus, only the maternal wild-type *MEA* allele is required for proper embryo, endosperm, and seed coat development (Kinoshita T, *et al.* (1999) *Plant Cell* 10:1945-52). These results reveal functions for plant polycomb proteins in the suppression of central cell proliferation and endosperm development (Kiyosue T, *et al. supra*).

Another gene product that controls seed development is *FIE*, also known as *FIE3* (see, e.g., copending U.S. patent application 09/071,838). The *FIE* protein is a homolog of the WD motif-containing Polycomb proteins from *Drosophila* and mammals (Ohad, N. *et al.* *Plant Cell* 11(3):407-16 (1999)). In *Drosophila*, these proteins function as repressors of homeotic genes. Loss of function mutations in the *FIE* gene result in endosperm phenotypes that are identical to *medea* loss of function mutations. A female gametophyte with a loss-of-function allele of *fie* undergoes replication of the central cell nucleus and initiates endosperm development without fertilization. These results suggest that the *FIE* Polycomb protein functions to suppress a critical aspect of early plant reproduction, namely, endosperm development, until fertilization occurs. Moreover, hypomethylation of *fie* mutants leads to the development of differentiated endosperm. Vinkenoog *et al.*, *Plant Cell* 12:2271-2282 (2000).

Control of the expression of genes that control egg and central cell differentiation, or those that control reproductive development, i.e. embryo, endosperm and seed coat, is useful in the production of plants with a range of desired traits. These and other advantages are provided by the present application.

## SUMMARY OF THE INVENTION

This invention provides isolated nucleic acids comprising a polynucleotide sequence, or its complement, encoding a DMT polypeptide comprising an amino acid sequence with at least 70% sequence identity to at least one of the following consensus sequences:

### DMT Domain A

KV<1>(I,l)D(D,p)(E,v)T<3>W<1>(L,v)L(M,l)(E,d)<0-  
2>D(K,e)<1>(K,t)<1>(K,a)(W,k)(W,l)<1>(E,k)ER<2>F<1>(G,t)R<1>(D,n)(S,l)FI(A,n)RM(  
H,r)<1>(V,l)QG(D,n)R<1>F<1>(P,q)WKGSVVDSV(I,v)GVFLTQN(V,t)D(H,y)(L,s)SS(S,n)A(F,  
y)M<1>(L,v)A(A,s)<1>FP

### DMT Domain B

W(D,n)<1>(L,f)R<5>E<3-  
6>D(S,t)<1>(D,n)(Y,w)<3>R<10>I<2>RG(M,q)(N,f)<2>L(A,s)<1>RI<2-  
12>FL<3>V<2>(H,n)G<1>IDLEWLR<2>(P,d)(P,s)(D,h)<1>(A,v)K<1>(Y,f)LL(S,e)(I,f)<1>G(  
L,l)GLKS(V,a)ECVRL<1>L(H,k)<2>AFFVDITNVGRI(A,c)VR(M,l)G(W,l)VPL(Q,e)PLP<2>(L,v)Q  
(L,m)H(L,q)L(E,f)<1>Y<1>(L,m)(E,d)(S,n)(I,v)QK(F,y)LWPRCLKL(D,p)Q<1>TLYELHY(Q,h  
) (L,m)ITFGK<0-2>FCTK<2>PNCNACPM(R,k)<0-2>EC(R,k)(H,y)(P,y)(A,s)SA<1>(A,v)<0-  
10>S(A,s)(R,k)<1>(A,l)L(P,e)<1>(P,t)

### DMT Domain C.

P(I,l)(I,v)E(E,f)P<1>(S,t)P<2-5>E<0-15>(D,a)IE(D,e)<4-  
23>(I,v)P<1>I<1>(L,f)(N,d)<8-17>(S,a)<1>(A,d)LV<8>(I,l)P<2-  
5>(K,r)(L,m)K<4>LRTEH<1>V(Y,f)(E,v)LPD<1>H<1>(L,i)L(E,k)<1>(D,e)D(P,i)<2>YLL(A,s  
)IW(T,q)P(G,d)(E,g)<6-8>(P,s)<3>C<6-  
10>(M,l)C<4>C<2>C<3>(R,k)E<5>(V,f)RGT(L,i)L<0-  
22>(L,v)FADH<1>(S,t)(S,r)<2>PI<3>(R,t)<3>(W,k)<1>L<1>(R,k)R<4>G(T,s)(S,t)<2>(S,t  
)I(F,c)(R,k)(G,l)L<1>(T,v)<2>I<2>(C,n)F(W,q)<1>G(F,y)(V,l)C(V,l)R<1>F(E,d)<3>(R,  
g)<1>P(R,k)<1>L<2>(R,h)LH<2>(A,v)SK

In some embodiments, the nucleic acids of the invention do not encode a polypeptide at least 40% identical to SEQ ID NO:2, or alternatively at least 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 99% or 100% to SEQ ID NO:2. In some embodiments, the DMT polypeptide comprises an amino acid sequence 100% identical to the above-listed consensus sequences.

In some embodiments, the DMT polypeptides are at least 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 97%, 98%, 99% or 100% identical to DMT domains A, B and/or C.

In one aspect, the invention provides DMt polypeptides capable of exhibiting at least one of the following biological activities:

- (a) glycosylase activity;
- (b) demethylation of polynucleotides;
- (c) DNA repair;
- (d) wherein expression of the polypeptide in a plant modulates organ identity;
- (e) wherein expression of the polypeptide in a plant modulates organ number;
- (f) wherein expression of the polypeptide in a plant modulate meristem stem and/or activity;
- (g) wherein enhanced expression of the polypeptide in a plant results in a delay in flowering time;
- (h) wherein introduction of the polypeptide into a cell results in modulation of methylation of chromosomal DNA in the cell;
- (i) wherein reduction of expression of the polypeptide in a plant results in modulation of endosperm development;
- (j) wherein expression of the polypeptide in an Arabidopsis leaf results in modulation of expression of the MEDEA gene.

In some aspects, the polypeptide comprises either a

- (i) basic region;
- (ii) nuclear localization signal;
- (iii) leucine zipper;
- (iv) helix-hairpin-helix structure;
- (v) glycine-proline rich loop with a terminal aspartic acid or
- (vi) helix that is capable of binding DNA.

In one aspect, the invention provides methods of modulating in a plant one or more of the following:

- (a) DNA repair;
- (b) wherein expression of the polypeptide in a plant modulates organ identity;
- (c) wherein expression of the polypeptide in a plant modulates organ number;

(d) wherein expression of the polypeptide in a plant modulate meristem stem and/or activity;

(e) wherein enhanced expression of the polypeptide in a plant results in a delay in flowering time;

(f) wherein introduction of the polypeptide into a cell results in modulation of methylation of chromosomal DNA in the cell;

(g) wherein reduction of expression of the polypeptide in a plant results in modulation of endosperm development;

(h) wherein expression of the polypeptide in an *Arabidopsis* leaf results in expression of the MEDEA gene,

wherein the method comprises:

(a) introducing into a plant cell a nucleic acid of claim 1; and

(b) generating conditions where the plant cell can transcribe the nucleic acid described above.

In some embodiments, the polypeptides comprise between 1500 and 2000 amino acids. In some aspects, the polypeptide has glycosylase activity. In some embodiments, introduction of the nucleic acid into a cell results in modulation of methylation of chromosomal DNA in the cell. In some embodiments, enhanced expression of the nucleic acids of the invention into a plant results in a delay in flowering time. In some embodiments, reduction of expression of a DMT polypeptide in a plant results in enhanced endosperm development. In addition, in some embodiments, expression of the nucleic acid of the invention in an *Arabidopsis* leaf results in expression of the MEDEA gene.

This invention provides isolated nucleic acids comprising a polynucleotide sequence, or its complement, encoding a DMT polypeptide exhibiting at least 60% sequence identity to SEQ ID NO:2 or exhibiting at least 70% sequence identity to at least one of DMT domain A, B, or C. For instance, the nucleic acid can encode the DMT polypeptide displayed in SEQ ID NO:2. In one aspect, the polynucleotide sequence comprises SEQ ID NO:5 or SEQ ID NO:1. In some aspects of the invention, the nucleic acid further comprises a promoter operably linked to the polynucleotide. In some embodiments, the promoter is constitutive. In other embodiments, the promoter is from a DMT gene. For example, the promoter can comprise a polynucleotide at least 70% identical to SEQ ID NO:3. In some aspects, the promoter comprises SEQ ID NO:3. In some aspects of this invention, the promoter further comprises a polynucleotide at least

70% identical to SEQ ID NO:4. For example, in some aspects the promoter comprises SEQ ID NO:4. In some aspects, the polynucleotide sequence is linked to the promoter in an antisense orientation.

The invention also provides an isolated nucleic acid molecule comprising a polynucleotide sequence exhibiting at least 60% sequence identity to SEQ ID NO:1.

The invention also provides an expression cassette comprising a promoter operably linked to a heterologous polynucleotide sequence, or complement thereof, encoding a DMT polypeptide exhibiting at least 60% sequence identity to SEQ ID NO:2. For instance, the nucleic acid can encode the DMT polypeptide displayed in SEQ ID NO:2. In some aspects, the polynucleotide sequence comprises SEQ ID NO:5 or SEQ ID NO:1. In some aspects of the invention, the nucleic acid further comprises a promoter operably linked to the polynucleotide. In some embodiments, the promoter is constitutive. In other embodiments, the promoter is from a DMT gene. For example, the promoter can comprise a polynucleotide at least 70% identical to SEQ ID NO:3. In some aspects, the promoter comprises SEQ ID NO:3. In some aspects of this invention, the promoter further comprises a polynucleotide at least 70% identical to SEQ ID NO:4. For example, in some aspects the promoter comprises SEQ ID NO:4. In some aspects, the polynucleotide sequence is linked to the promoter in an antisense orientation.

The invention also provides an expression cassette for the expression of a heterologous polynucleotide in a plant cell. In some aspects, the expression cassette comprises a promoter polynucleotide at least 70% identical to SEQ ID NO:3 that is operably linked to a heterologous polynucleotide. In some aspects, the promoter comprises SEQ ID NO:3. In some aspects, the promoter further comprises a polynucleotide at least 70% identical to SEQ ID NO:4. For instance, in some embodiments, the promoter comprises SEQ ID NO:4. In some aspects, the promoter further comprises a polynucleotide at least 70% identical to SEQ ID NO:6. In some aspects, the promoter comprises SEQ ID NO:6.

The present invention also provides a host cell comprising an exogenous polynucleotide sequence comprising a polynucleotide sequence, or complement thereof, encoding a DMT polypeptide exhibiting at least 60% sequence identity to SEQ ID NO:2 or exhibiting at least 70% sequence identity to at least one of DMT domain A, B, or C. In some aspects of the invention, the nucleic acid further comprises a promoter operably linked to the polynucleotide sequence. In some aspects, the promoter is constitutive. In some aspects, the promoter comprises a polynucleotide at least 70% identical to SEQ ID

NO:3. The promoter, for instance, can comprise SEQ ID NO:3. In some aspects, the promoter further comprises a polynucleotide at least 70% identical to SEQ ID NO:4. For instance, in some embodiments, the promoter comprises SEQ ID NO:4. In some aspects, the promoter is operably linked to the exogenous polynucleotide sequence in an antisense orientation.

The present invention also provides an isolated polypeptide comprising an amino acid sequence at least 60% identical to SEQ ID NO:2 or an amino acid sequence at least 70% sequence identical to at least one of DMT domain A, B, or C and capable of exhibiting at least one biological activity of the polypeptide displayed in SEQ ID NO:2, or fragment thereof. The present invention also provides for an antibody capable of binding such polypeptides.

The present invention also provides a method of introducing an isolated nucleic acid into a host cell comprising, (a) providing an isolated nucleic acid or its complement, encoding a DMT polypeptide exhibiting at least 60% sequence identity to SEQ ID NO:2 or exhibiting at least 70% sequence identity to at least one of DMT domain A, B, or C and (b) contacting the nucleic acid with the host cell under conditions that permit insertion of the nucleic acid into the host cell.

The present invention also provides a method of modulating transcription, comprising introducing into a host cell an expression cassette comprising a promoter operably linked to a heterologous DMT polynucleotide, the heterologous DMT polynucleotide encoding a DMT polypeptide at least 60% identical to SEQ ID NO:2 or at least 70% sequence identical to at least one of DMT domain A, B, or C, and detecting a host cell with modulated transcription. In some aspects of the invention, the heterologous DMT polynucleotide encodes SEQ ID NO:2. In some aspect, the polynucleotide sequence comprises SEQ ID NO:5 or SEQ ID NO:1. In some aspects, the expression cassette is introduced into a host cell by *Agrobacterium*. In some aspects, the expression cassette is introduced by a sexual cross. In some aspects of the method of the invention, modulating transcription results in the modulation of endosperm development in a plant. In some aspects, endosperm development is enhanced. In other aspects, endosperm development is decreased. In some aspects of the methods of the invention, the promoter is operably linked to the DMT polynucleotide in an antisense orientation.

The present invention also provides a method of detecting a nucleic acid in a sample, comprising (a) providing an isolated nucleic acid molecule comprising a polynucleotide sequence, or its complement, encoding a DMT polypeptide exhibiting at



least 60% sequence identity to SEQ ID NO:2 or exhibiting at least 70% sequence identity to at least one of DMT domain A, B, or C, (b) contacting the isolated nucleic acid molecule with a sample under conditions that permit a comparison of the sequence of the isolated nucleic acid molecule with the sequence of DNA in the sample, and (c) analyzing the result of the comparison. In some aspects of the method, the isolated nucleic acid molecule and the sample are contacted under conditions that permit the formation of a duplex between complementary nucleic acid sequences.

The present invention also provides a transgenic plant cell or transgenic plant comprising a polynucleotide sequence, or its complement, encoding a DMT polypeptide exhibiting at least 60% sequence identity to SEQ ID NO:2 or exhibiting at least 70% sequence identity to at least one of DMT domain A, B, or C. For instance, the nucleic acid can encode the DMT polypeptide displayed in SEQ ID NO:2. In one aspect, the polynucleotide sequence comprises SEQ ID NO:5 or SEQ ID NO:1. In some aspects of the invention, the nucleic acid further comprises a promoter operably linked to the polynucleotide. In some embodiments, the promoter is constitutive. In other embodiments, the promoter comprises a polynucleotide at least 70% identical to SEQ ID NO:3. In some aspects, the promoter comprises SEQ ID NO:3. In some aspects of this invention, the promoter further comprises a polynucleotide at least 70% identical to SEQ ID NO:4. For example, in some aspects the promoter comprises SEQ ID NO:4. In some aspects, the polynucleotide sequence is linked to the promoter in an antisense orientation. The present invention also provides a plant that is regenerated from a plant cell as described above.

The present invention also provides an expression cassette for the expression of a heterologous polynucleotide in a plant cell, wherein the expression cassette comprises a promoter at least 70% identical to SEQ ID NO:3 and the promoter is operably linked to a heterologous polynucleotide. In some embodiments, the promoter comprises a polynucleotide at least 70% identical to SEQ ID NO:4 and/or SEQ ID NO:6. In some embodiments, the promoter specifically directs expression of the heterologous polynucleotide in a female gametophyte when the expression cassette is introduced into a plant.

#### DEFINITIONS

The phrase "nucleic acid sequence" refers to a single or double-stranded polymer of deoxyribonucleotide or ribonucleotide bases read from the 5' to the 3' end. It

includes chromosomal DNA, self-replicating plasmids, infectious polymers of DNA or RNA and DNA or RNA that performs a primarily structural role.

A "promoter" is defined as an array of nucleic acid control sequences that direct transcription of an operably linked nucleic acid. As used herein, a "plant promoter" is a promoter that functions in plants. Promoters include necessary nucleic acid sequences near the start site of transcription, such as, in the case of a polymerase II type promoter, a TATA element. A promoter also optionally includes distal enhancer or repressor elements, which can be located as much as several thousand base pairs from the start site of transcription. A "constitutive" promoter is a promoter that is active under most environmental and developmental conditions. An "inducible" promoter is a promoter that is active under environmental or developmental regulation. The term "operably linked" refers to a functional linkage between a nucleic acid expression control sequence (such as a promoter, or array of transcription factor binding sites) and a second nucleic acid sequence, wherein the expression control sequence directs transcription of the nucleic acid corresponding to the second sequence.

The term "plant" includes whole plants, plant organs (e.g., leaves, stems, flowers, roots, etc.), seeds and plant cells and progeny of same. The class of plants which can be used in the method of the invention is generally as broad as the class of flowering plants amenable to transformation techniques, including angiosperms (monocotyledonous and dicotyledonous plants), as well as gymnosperms. It includes plants of a variety of ploidy levels, including polyploid, diploid, haploid and hemizygous.

A polynucleotide sequence is "heterologous to" an organism or a second polynucleotide sequence if it originates from a foreign species, or, if from the same species, is modified from its original form. For example, a promoter operably linked to a heterologous coding sequence refers to a coding sequence from a species different from that from which the promoter was derived, or, if from the same species, a coding sequence which is different from any naturally occurring allelic variants.

A polynucleotide "exogenous to" an individual plant is a polynucleotide which is introduced into the plant, or a predecessor generation of the plant, by any means other than by a sexual cross. Examples of means by which this can be accomplished are described below, and include Agrobacterium-mediated transformation, biolistic methods, electroporation, in planta techniques, and the like. "Exogenous," as referred to within, is any polynucleotide, polypeptide or protein sequence, whether chimeric or not, that is initially or subsequently introduced into the genome of an individual host cell or the

organism regenerated from said host cell by any means other than by a sexual cross. Examples of means by which this can be accomplished are described below, and include Agrobacterium-mediated transformation (of dicots - e.g. Salomon et al. EMBO J. 3:141 (1984); Herrera-Estrella et al. EMBO J. 2:987 (1983); of monocots, representative papers are those by Escudero et al., Plant J. 10:355 (1996), Ishida et al., Nature Biotechnology 14:745 (1996), May et al., Bio/Technology 13:486 (1995)), biolistic methods (Armaleo et al., Current Genetics 17:97 1990)), electroporation, in planta techniques, and the like. Such a plant containing the exogenous nucleic acid is referred to here as a T0 for the primary transgenic plant and T1 for the first generation. The term "exogenous" as used herein is also intended to encompass inserting a naturally found element into a non-naturally found location.

The phrase "host cell" refers to a cell from any organism. Preferred host cells are derived from plants, bacteria, yeast, fungi, insects or other animals, including humans. Methods for introducing polynucleotide sequences into various types of host cells are well known in the art.

The "biological activity of a polypeptide" refers to any molecular activity or phenotype that is caused by the polypeptide. For example, the ability to transfer a phosphate to a substrate or the ability to bind a specific DNA sequence is a biological activity. One biological activity of DMT is glycosylase activity, i.e., cleavage of the nucleotide base from the nucleotide sugar). Another biological activity of DMT is to demethylate nucleotides (e.g., DMT has 5'-methylcytosine glycosylase activity). In addition, DMT has the ability to modulate endosperm production, as described herein, and to modulate flowering time in plants. For example, when DMT expression or DMT activity is increased in a plant, the flowering time of the plant is delayed. Moreover, expression of a DMT polypeptide in a plant tissue (e.g., a leaf) that does not typically express the *MEDEA* gene (Grossniklaus U, et al., Science 280(5362):446-50 (1998)) results in the expression of *MEDEA*.

Additional biological activities of DMT polypeptides include: nuclear localization (e.g., as localized by amino acids 43-78 of SEQ ID NO:2); the ability to modulate plant organ size and/or number; the ability to modulate meristem size and/or activity; and to perform DNA repair, including nucleotide methylation or demethylation and/or repair and/or removal of mis-matched nucleotides from DNA.

An "expression cassette" refers to a nucleic acid construct, which when introduced into a host cell, results in transcription and/or translation of an RNA or

polypeptide, respectively. Antisense or sense constructs that are not or cannot be translated are expressly included by this definition.

A "DMT nucleic acid" or "DMT polynucleotide sequence" of the invention is a subsequence or full length polynucleotide sequence of a gene which encodes a polypeptide involved in control of reproductive development and which, when the maternal allele is mutated or when DMT activity is reduced or eliminated in a maternal tissue or plant, allows for increased production of the endosperm and/or abortion of the embryo. In addition, overexpression of DMT in plants results in delayed time to flowering. Moreover, DMT is necessary and sufficient for expression of *MEDEA* in a plant cell. An exemplary nucleic acid of the invention is the *Arabidopsis* DMT sequence (SEQ ID NO:1). Additional DMT nucleic acid sequences from a variety of plant species are also provided (e.g., SEQ ID NOs: 7-70). DMT polynucleotides are defined by their ability to hybridize under defined conditions to the exemplified nucleic acids or PCR products derived from them. A DMT polynucleotide is typically at least about 30-40 nucleotides to about 7000, usually less than about 10,000 nucleotides in length. More preferably, DMT polynucleotides contain a coding sequence of from about 100 to about 5500 nucleotides, often from about 500 to about 3600 nucleotides in length. A DMT polypeptide is typically at least 500 amino acids, typically at least 1000 amino acids, more typically at least 1500 amino acids. In some embodiments, a DMT polypeptide comprises fewer than 2000 amino acids, more typically fewer than 3000 amino acid and still more typically fewer than 5000 or 7500 amino acid in length.

As described below, DMT nucleic acid sequences encode polypeptides with substantial identity to at least one of following the consensus sequences:

#### DMT Domain A

KV<1>(I,1)D(D,p)(E,v)T<3>W<1>(L,v)L(M,1)(E,d)<0-  
2>D(K,e)<1>(K,t)<1>(K,a)(W,k)(W,l)<1>(E,k)ER<2>P<1>(G,t)R<1>(D,n)(S,l)FI(A,n)RM(  
H,r)<1>(V,l)QG(D,n)R<1>F<1>(P,q)WKGSVVDSV(I,v)GVFLTQN(V,t)D(H,y)(L,s)SS(S,n)A(F,  
y)M<1>(L,v)A(A,s)<1>FP

#### DMT Domain B

W(D,n)<1>(L,f)R<5>E<3-  
6>D(S,t)<1>(D,n)(Y,w)<3>R<10>I<2>RG(M,q)(N,f)<2>L(A,s)<1>RI<2-  
12>FL<3>V<2>(H,n)G<1>IDLEWLR<2>(P,d)(P,s)(D,h)<1>(A,v)K<1>(Y,f)LL(S,e)(I,f)<1>G(  
L,i)GLKS(V,a)ECVRL<1>L(H,k)<2>AFPVDITNVGRI(A,c)VR(M,1)G(W,l)VPL(Q,e)PLP<2>(L,v)Q  
(L,m)H(L,q)L(E,f)<1>YP<1>(L,m)(E,d)(S,n)(I,v)QK(F,y)LWPRLCKL(D,p)Q<1>TYELHY(Q,h

) (L,m) ITPGK<0-2>FCTK<2>PNCNACPM(R,k)<0-2>EC(R,k) (H,y) (F,y) (A,s) SA<1>(A,v)<0-10>S(A,s) (R,k)<1>(A,l) L(P,e)<1>(P,t)

# DMT Domain C.

- 5 P(I,1) (I,v) E(E,f) P<1>(S,t) P<2-5>E<0-15>(D,a) IE(D,e)<4-23>(I,v) P<1>I<1>(L,f) (N,d)<8-17>(S,a)<1>(A,d) LV<8>(I,1) P<2-5>(K,r) (L,m) K<4>LRTEH<1>V(Y,f) (E,v) LPD<1>H<1>(L,i) L(E,k)<1>(D,e) D(P,i)<2>YLL(A,s) IW(T,q) P(G,d) (E,g)<6-8>(P,s)<3>C<6-10>(M,l) C<4>C<2>C<3>(R,k) E<5>(V,f) RGT(L,i) L<0-22>(L,v) PADH<1>(S,t) (S,r)<2>PI<3>(R,t)<3>(W,k)<1>L<1>(R,k) R<4>G(T,s) (S,t)<2>(S,t) I(F,c) (R,k) (G,l) L<1>(T,v)<2>I<2>(C,n) F(W,q)<1>G(F,y) (V,l) C(V,l) R<1>F(E,d)<3>(R,g)<1>P(R,k)<1>L<2>(R,h) LH<2>(A,v) SK

In addition, the following consensus sequence spanning all three domains

- 15 were identified:  
 <9-14>(T,q) (A,i) (S,k) (I,1)<3>(A,r) (S,k)<1>(G,m)<2>(S,r) (P,k)<2>(K,f)<2>(E,l) K<0-1>K<0-3>(P,r)<2>(P,r)<1>(K,r) (K,r) (G,d) (R,k)<1>(G,v)<1>(K,g)<3-5>(P,s) (P,k)<3>(S,n)<1>(I,1)<0-2>(Q,d)<9>(P,q)<4>(K,a) (P,s)<14-16>(P,a)<4>L<0-10>D<1>(I,1)<0-4>(L,n)<12-46>(K,d)<2-7>(P,a) KV<1>(I,1) D(D,p) (E,v) T<3>W<1>(L,v) L(M,l) (E,d)<0-2>D(K,e)<1>(K,t)<1>(K,a) (W,k) (W,l)<1>(E,k) ER<2>P<1>(G,t) R<1>(D,n) (S,l) FI(A,r) NM(H,r)<1>(V,l) QG(D,n) R<1>P<1>(P,q) WKGSVVDV(I,v) GVFLTQN(V,t) D(H,y) (L,s) SS(S,n) A(F,y) M<1>(L,v) A(A,s)<1>FP<0-16>(P,v)<6-15>(S,h)<3>(E,d)<10-24>(S,t)<1>(S,e)<6>(K,n)<8-55>(E,i)<8-9>(I,v)<1>(N,s)<1-4>(E,d)<1>(E,s)<4>(Q,l)<0-11>(D,h)<1>(P,m)<5>(Q,n)<0-3>(G,e)<2>(G,d) S<1>(K,d)<7-11>(T,m)<2>(V,l)<3>(S,q)<6-10>(S,e)<2-3>(S,v)<19-25>(T,s)<16-28>(R,s)<2-6>(T,p)<5>(P,k)<10>(Q,e)<4>(D,s)<1-4>(S,r)<5>(D,p)<3>(N,d)<3>(P,y)<2>(F,s)<1>(R,k)<1>(G,s)<1>(S,a) (V,r) (P,e)<3>(T,s)<3-6>(I,1)<3>(P,e)<1>E<3-5>(L,q)<1>(G,c)<1>(S,h) (S,n)<1>(V,q)<1>(E,d)<3>T(Q,e)<1-2>(N,g)<3>(E,n)<20-30>(N,a) (P,g)<1-6>(S,l)<25-46>(Q,d) W(D,n)<1>(L,f) R<5>E<3-6>D(S,t)<1>(D,n) (Y,w)<3>R<10>I<2>RG(M,q) (N,f)<2>L(A,s)<1>RI<2-12>FL<3>V<2>(H,n) G<1>IDLEWL<2>(P,d) (P,s) (D,h)<1>(A,v) K<1>(Y,f) LL(S,e) (I,f)<1>G(L,i) GLKS(V,a) ECVRL<1>L(H,k)<2>AFFVDITNVGRI(A,c) VR(M,l) G(W,l) VPL(Q,e) PLP<2>(L,v) Q(L,m) H(L,q) L(E,f)<1>YP<1>(L,m) (E,d) (S,n) (I,v) QK(F,y) LWPRCLKL(D,p) Q<1>TYELHY(Q,h) (L,m) ITPGK<0-2>FCTK<2>PNCNACPM(R,k)<0-2>EC(R,k) (H,y) (F,y) (A,s) SA<1>(A,v)<0-10>S(A,s) (R,k)<1>(A,l) L(P,e)<1>(P,t) (E,q)<7-16>P(I,1) (I,v) E(E,f) P<1>(S,t) P<2-5>E<0-15>(D,a) IE(D,e)<4-23>(I,v) P<1>I<1>(L,f) (N,d)<8-17>(S,a)<1>(A,d) LV<8>(I,1) P<2-5>(K,r) (L,m) K<4>LRTEH<1>V(Y,f) (E,v) LPD<1>H<1>(L,i) L(E,k)<1>(D,e) D(P,i)<2>YLL(A,s) IW(T,q) P(G,d) (E,g)<6-8>(P,s)<3>C<6-10>(M,l) C<4>C<2>C<3>(R,k) E<5>(V,f) RGT(L,i) L<0-22>(L,v) PADH<1>(S,t) (S,r)<2>PI<3>(R,t)<3>(W,k)<1>L<1>(R,k) R<4>G(T,s) (S,t)<2>(S,t) I(F,c) (R,k) (G,l) L<1>(T,v)<2>I<2>(C,n) F(W,q)<1>G(F,y) (V,l) C(V,l) R<1>F(E,d)<3>(R,g)<1>P(R,k)<1>L<2>(R,h) LH<2>(A,v) SK

) I (F, c) (R, k) (G, l) L<1> (T, v) <2> I<2> (C, n) F (W, q) <1> G (F, y) (V, l) C (V, l) R<1> F (E, d) <3> (R, g) <1> P (R, k) <1> L<2> (R, h) L H<2> (A, v) S K

DMT domain A corresponds to amino acid positions 697 through 796 of SEQ ID NO:2. DMT domain B corresponds to amino acid positions 1192 through 1404 of SEQ ID NO:2. DMT domain C corresponds to amino acid positions 1452 through 1722 of SEQ ID NO:2. The consensus sequence provides amino acid sequences by position using single letter amino acid abbreviations. Numbers in carrots (“<” or “>”) refer to amino acid positions where there is no consensus and which therefore, can be any amino acid. Amino acid abbreviations in parentheses indicate alternative amino acids at the same position. Capitalized letters refer to predominant consensus amino acids and lower case letters refer to amino acids that are commonly found in DMT sequences, but are not predominant. Thus, it is a simple matter to identify whether any particular nucleic acid sequence is a DMT nucleic acid and/or encodes a DMT polypeptide.

The structure of full-length DMT polypeptides comprises the following domains and regions. These regions are generally described with reference to SEQ ID NO:2. First, as described above, domain B DMT polypeptides can comprise a bipartite nuclear localization signal (e.g., amino acid positions 43-60 and 61-78 in SEQ ID NO:2) comprised of basic amino acids. Amino acids 36-91 are homologous to human G/T mismatch-specific thymine DNA glycosylase (Genbank accession number AAC50540.1), which has 5-methylcytosine glycosylase activity (Zhu *et al.*, *Nuc. Acids Res.* 28:4157-4165 (2000)). DMT polypeptides also contain a leucine zipper sequence (e.g., positions 1330-1351 of SEQ ID NO:2), that can be involved in protein-protein interactions as well as DNA binding. In addition, the amino portion of the DMT polypeptide (amino acids 43-78) is generally basic, similar to histone H1. Thus, without intending to limit the scope of the invention, it is believed this basic portion of DMT facilitates interactions with DNA and/or chromatic proteins.

In addition, amino acids 1-800 is related to the beta subunit of bacterial DNA-dependent RNA polymerases. Without intending to limit the scope of the invention, it is believed the RNA polymerase-like domain facilitates interaction of DMT with DNA.

Amino acids 1167-1368 is related to proteins in the HhH-GPD superfamily. Amino acids 1,271 to 1,304 correspond to the conserved HhH-GPD motif. The corresponding DMT sequence is DKAKDYLLSIRGLGLKSVECVRLTLHNLAFPVD. Secondary structure prediction

(Jpred program) indicates that DMT has two alpha-helices (1,271 - 1,279 and 1,286 to 1,295) that correspond to the conserved alphaK and alphaL helices in the HhH-GPD motif of the crystallized hOGG1 DNA repair protein (Bruner et al *Nature* 403:859-866 (2000)). In between the two helices (1280 to 1285), is a hairpin with conserved glycines (G1282 and G1284). Amino acids 1286 to 1295 are related to the alphaL helix of hOGG1, which contacts the DNA backbone (Bruner et al *Nature* 403:859-866 (2000)). Thus, without intending to limit the scope of the invention, it is believed this region of DMT contacts the DNA. The catalytic lysine (K1286) and aspartic acid (D1304) residues are conserved in the HhH-GPD motif of DMT. Without intending to limit the scope of the invention, by analogy to hOGG1, K1286 is predicted to displace the modified base and to promote conjugate elimination of the 3'-phosphodiester bond. Without intending to limit the scope of the invention, by analogy to hOGG1, D1304 is believed to assist the reaction by transferring protons to and from K1286.

DMT nucleic acids are a new class of plant regulatory genes that encode polypeptides with sequence identity to members of the endonuclease III genes found in a diverse collection of organisms. Endonuclease III is implicated in various DNA repair reactions. Thus proteins related to endonuclease III are likely to have a chromosomal function. DMT (SEQ ID NO:1) is most related to endonuclease III from *Deinococcus radiodurans* Genbank Accession No. AE002073 (see, e.g., White, O. et al. *Science* 286:1571-1577 (1999)). DMT polypeptides have glycosylase activity (i.e., the capability to cleave the base portion of a nucleotide from the sugar portion). More particularly, DMT polypeptides have demethylase activity, and in more preferred embodiments, have 5-methylcytosine glycosylase activity. Demethylation activity can be assayed *in vivo* by expressing a candidate polypeptide in the nucleus of a cell and then assaying for a change in methylation of the cell's DNA. See, e.g., Vong, et al., *Science* 260:1926-1928 (1993). Changes in chromosomal methylation can be measured by comparing the ability of methylation sensitive and insensitive endonucleases to cleave DNA from a cell expressing a polypeptide suspected of having demethylase or methylase activity. Alternatively, bisulfate sequencing can be used to identify which base pairs are methylated in a DNA sequence. For a discussion of both methods, see Soppe et al., *Molec. Cell.* 6:791-802 (2000). *In vitro* assays to measure demethylase activity using labeled substrates are also known to those of skill in the art. See, e.g., Vhu et al., *Proc. Natl. Acad. Sci. USA* 97:5135-5139 (2000).

09840743.042301

In the case of both expression of transgenes and inhibition of endogenous genes (e.g., by antisense, or sense suppression) one of skill will recognize that the inserted polynucleotide sequence need not be identical, but may be only "substantially identical" to a sequence of the gene from which it was derived. As explained below, these

5 substantially identical variants are specifically covered by the term *DMT* nucleic acid.

In the case where the inserted polynucleotide sequence is transcribed and translated to produce a functional polypeptide, one of skill will recognize that because of codon degeneracy a number of polynucleotide sequences will encode the same polypeptide. These variants are specifically covered by the terms "*DMT* nucleic acid". In addition, the term specifically includes those sequences substantially identical (determined as described below) with a *DMT* polynucleotide sequence disclosed here and that encode polypeptides that are either mutants of wild type *DMT* polypeptides or retain the function of the *DMT* polypeptide (e.g., resulting from conservative substitutions of amino acids in the *DMT* polypeptide). In addition, variants can be those that encode dominant negative mutants as described below.

Two nucleic acid sequences or polypeptides are said to be "identical" if the sequence of nucleotides or amino acid residues, respectively, in the two sequences is the same when aligned for maximum correspondence as described below. The terms "identical" or percent "identity," in the context of two or more nucleic acids or polypeptide sequences, refer to two or more sequences or subsequences that are the same or have a specified percentage of amino acid residues or nucleotides that are the same, when compared and aligned for maximum correspondence over a comparison window, as measured using one of the following sequence comparison algorithms or by manual alignment and visual inspection. When percentage of sequence identity is used in reference to proteins or peptides, it is recognized that residue positions that are not identical often differ by conservative amino acid substitutions, where amino acids residues are substituted for other amino acid residues with similar chemical properties (e.g., charge or hydrophobicity) and therefore do not change the functional properties of the molecule. Where sequences differ in conservative substitutions, the percent sequence identity may be adjusted upwards to correct for the conservative nature of the substitution. Means for making this adjustment are well known to those of skill in the art. Typically this involves scoring a conservative substitution as a partial rather than a full mismatch, thereby increasing the percentage sequence identity. Thus, for example, where an identical amino acid is given a score of 1 and a non-conservative substitution is given a



score of zero, a conservative substitution is given a score between zero and 1. The scoring of conservative substitutions is calculated according to, e.g., the algorithm of Meyers & Miller, *Computer Applic. Biol. Sci.* 4:11-17 (1988) e.g., as implemented in the program PC/GENE (Intelligenetics, Mountain View, California, USA).

5           The phrase "substantially identical," in the context of two nucleic acids or polypeptides, refers to a sequence or subsequence that has at least 40% sequence identity with a reference sequence. Alternatively, percent identity can be any integer from 40% to 100%. More preferred embodiments include at least: 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, or 99% compared to a reference sequence  
10       using the programs described herein; preferably BLAST using standard parameters, as described below. This definition also refers to the complement of a test sequence, when the test sequence has substantial identity to a reference sequence.

          For sequence comparison, typically one sequence acts as a reference sequence, to which test sequences are compared. When using a sequence comparison  
15       algorithm, test and reference sequences are entered into a computer, subsequence coordinates are designated, if necessary, and sequence algorithm program parameters are designated. Default program parameters can be used, or alternative parameters can be designated. The sequence comparison algorithm then calculates the percent sequence identities for the test sequences relative to the reference sequence, based on the program  
20       parameters.

          A "comparison window", as used herein, includes reference to a segment of any one of the number of contiguous positions selected from the group consisting of from 20 to 600, usually about 50 to about 200, more usually about 100 to about 150 in which a sequence may be compared to a reference sequence of the same number of  
25       contiguous positions after the two sequences are optimally aligned. Methods of alignment of sequences for comparison are well-known in the art. Optimal alignment of sequences for comparison can be conducted, e.g., by the local homology algorithm of Smith & Waterman, *Adv. Appl. Math.* 2:482 (1981), by the homology alignment algorithm of Needleman & Wunsch, *J. Mol. Biol.* 48:443 (1970), by the search for  
30       similarity method of Pearson & Lipman, *Proc. Nat'l. Acad. Sci. USA* 85:2444 (1988), by computerized implementations of these algorithms (GAP, BESTFIT, FASTA, and TFASTA in the Wisconsin Genetics Software Package, Genetics Computer Group, 575 Science Dr., Madison, WI), or by manual alignment and visual inspection.

09840743-042301

One example of a useful algorithm is PILEUP. PILEUP creates a multiple sequence alignment from a group of related sequences using progressive, pairwise alignments to show relationship and percent sequence identity. It also plots a tree or dendrogram showing the clustering relationships used to create the alignment. PILEUP uses a simplification of the progressive alignment method of Feng & Doolittle, *J. Mol. Evol.* 35:351-360 (1987). The method used is similar to the method described by Higgins & Sharp, *CABIOS* 5:151-153 (1989). The program can align up to 300 sequences, each of a maximum length of 5,000 nucleotides or amino acids. The multiple alignment procedure begins with the pairwise alignment of the two most similar sequences, producing a cluster of two aligned sequences. This cluster is then aligned to the next most related sequence or cluster of aligned sequences. Two clusters of sequences are aligned by a simple extension of the pairwise alignment of two individual sequences. The final alignment is achieved by a series of progressive, pairwise alignments. The program is run by designating specific sequences and their amino acid or nucleotide coordinates for regions of sequence comparison and by designating the program parameters. For example, a reference sequence can be compared to other test sequences to determine the percent sequence identity relationship using the following parameters: default gap weight (3.00), default gap length weight (0.10), and weighted end gaps.

Another example of algorithm that is suitable for determining percent sequence identity and sequence similarity is the BLAST algorithm, which is described in Altschul *et al.*, *J. Mol. Biol.* 215:403-410 (1990). Software for performing BLAST analyses is publicly available through the National Center for Biotechnology Information (<http://www.ncbi.nlm.nih.gov/>). This algorithm involves first identifying high scoring sequence pairs (HSPs) by identifying short words of length W in the query sequence, which either match or satisfy some positive-valued threshold score T when aligned with a word of the same length in a database sequence. T is referred to as the neighborhood word score threshold (Altschul *et al.*, *supra*). These initial neighborhood word hits act as seeds for initiating searches to find longer HSPs containing them. The word hits are extended in both directions along each sequence for as far as the cumulative alignment score can be increased. Extension of the word hits in each direction are halted when: the cumulative alignment score falls off by the quantity X from its maximum achieved value; the cumulative score goes to zero or below, due to the accumulation of one or more negative-scoring residue alignments; or the end of either sequence is reached. The BLAST algorithm parameters W, T, and X determine the sensitivity and speed of the

alignment. The BLAST program uses as defaults a wordlength (W) of 11, the BLOSUM62 scoring matrix (see Henikoff & Henikoff, *Proc. Natl. Acad. Sci. USA* 89:10915 (1989)) alignments (B) of 50, expectation (E) of 10, M=5, N=4, and a comparison of both strands.

5       The BLAST algorithm also performs a statistical analysis of the similarity between two sequences (see, e.g., Karlin & Altschul, *Proc. Nat'l. Acad. Sci. USA* 90:5873-5787 (1993)). One measure of similarity provided by the BLAST algorithm is the smallest sum probability (P(N)), which provides an indication of the probability by which a match between two nucleotide or amino acid sequences would occur by chance.

10   For example, a nucleic acid is considered similar to a reference sequence if the smallest sum probability in a comparison of the test nucleic acid to the reference nucleic acid is less than about 0.2, more preferably less than about 0.01, and most preferably less than about 0.001.

      "Conservatively modified variants" applies to both amino acid and nucleic acid sequences. With respect to particular nucleic acid sequences, conservatively modified variants refers to those nucleic acids which encode identical or essentially identical amino acid sequences, or where the nucleic acid does not encode an amino acid sequence, to essentially identical sequences. Because of the degeneracy of the genetic code, a large number of functionally identical nucleic acids encode any given protein.

15   For instance, the codons GCA, GCC, GCG and GCU all encode the amino acid alanine. Thus, at every position where an alanine is specified by a codon, the codon can be altered to any of the corresponding codons described without altering the encoded polypeptide. Such nucleic acid variations are "silent variations," which are one species of conservatively modified variations. Every nucleic acid sequence herein which encodes a

20   polypeptide also describes every possible silent variation of the nucleic acid. One of skill will recognize that each codon in a nucleic acid (except AUG, which is ordinarily the only codon for methionine) can be modified to yield a functionally identical molecule. Accordingly, each silent variation of a nucleic acid which encodes a polypeptide is implicit in each described sequence.

25   As to amino acid sequences, one of skill will recognize that individual substitutions, deletions or additions to a nucleic acid, peptide, polypeptide, or protein sequence which alters, adds or deletes a single amino acid or a small percentage of amino acids in the encoded sequence is a "conservatively modified variant" where the alteration results in the substitution of an amino acid with a chemically similar amino acid.

Conservative substitution tables providing functionally similar amino acids are well known in the art.

The following six groups each contain amino acids that are conservative substitutions for one another:

- 1) Alanine (A), Serine (S), Threonine (T);
- 2) Aspartic acid (D), Glutamic acid (E);
- 3) Asparagine (N), Glutamine (Q);
- 4) Arginine (R), Lysine (K);
- 5) Isoleucine (I), Leucine (L), Methionine (M), Valine (V); and
- 6) Phenylalanine (F), Tyrosine (Y), Tryptophan (W).

(see, e.g., Creighton, *Proteins* (1984)).

An indication that two nucleic acid sequences or polypeptides are substantially identical is that the polypeptide encoded by the first nucleic acid is immunologically cross reactive with the antibodies raised against the polypeptide encoded by the second nucleic acid. Thus, a polypeptide is typically substantially identical to a second polypeptide, for example, where the two peptides differ only by conservative substitutions. Another indication that two nucleic acid sequences are substantially identical is that the two molecules or their complements hybridize to each other under stringent conditions, as described below.

The phrase "selectively (or specifically) hybridizes to" refers to the binding, duplexing, or hybridizing of a molecule only to a particular nucleotide sequence under stringent hybridization conditions when that sequence is present in a complex mixture (e.g., total cellular or library DNA or RNA).

The phrase "stringent hybridization conditions" refers to conditions under which a probe will hybridize to its target subsequence, typically in a complex mixture of nucleic acid, but to no other sequences. Stringent conditions are sequence-dependent and will be different in different circumstances. Longer sequences hybridize specifically at higher temperatures. An extensive guide to the hybridization of nucleic acids is found in Tijssen, *Techniques in Biochemistry and Molecular Biology--Hybridization with Nucleic Probes*, "Overview of principles of hybridization and the strategy of nucleic acid assays" (1993). Generally, highly stringent conditions are selected to be about 5-10°C lower than the thermal melting point ( $T_m$ ) for the specific sequence at a defined ionic strength pH. Low stringency conditions are generally selected to be about 15-30°C below the  $T_m$ . The

$T_m$  is the temperature (under defined ionic strength, pH, and nucleic concentration) at which 50% of the probes complementary to the target hybridize to the target sequence at equilibrium (as the target sequences are present in excess, at  $T_m$ , 50% of the probes are occupied at equilibrium). Stringent conditions will be those in which the salt

concentration is less than about 1.0 M sodium ion, typically about 0.01 to 1.0 M sodium ion concentration (or other salts) at pH 7.0 to 8.3 and the temperature is at least about 30°C for short probes (e.g., 10 to 50 nucleotides) and at least about 55°C, sometimes 60°C, and sometimes 65°C for long probes (e.g., greater than 50 nucleotides). Stringent conditions may also be achieved with the addition of destabilizing agents such as formamide. For selective or specific hybridization, a positive signal is at least two times background, preferably 10 time background hybridization.

Nucleic acids that do not hybridize to each other under stringent conditions are still substantially identical if the polypeptides which they encode are substantially identical. This occurs, for example, when a copy of a nucleic acid is created using the maximum codon degeneracy permitted by the genetic code. In such cases, the nucleic acids typically hybridize under moderately stringent hybridization conditions.

In the present invention, genomic DNA or cDNA comprising *DMT* nucleic acids of the invention can be identified in standard Southern blots under stringent conditions using the nucleic acid sequences disclosed here. For the purposes of this disclosure, suitable stringent conditions for such hybridizations are those which include a hybridization in a buffer of 40% formamide, 1 M NaCl, 1% SDS at 37°C, and at least one wash in 0.2X SSC at a temperature of at least about 50°C, usually about 55°C to about 60°C and sometimes 65°C, for 20 minutes, or equivalent conditions. A positive hybridization is at least twice background. Those of ordinary skill will readily recognize that alternative hybridization and wash conditions can be utilized to provide conditions of similar stringency.

A further indication that two polynucleotides are substantially identical is if the reference sequence, amplified by a pair of oligonucleotide primers, can then be used as a probe under stringent hybridization conditions to isolate the test sequence from a cDNA or genomic library, or to identify the test sequence in, e.g., a northern or Southern blot.

## DETAILED DESCRIPTION

This invention provides molecular strategies for controlling plant development, including methylation of chromosomal DNA, endosperm development and flowering time.

5       Reproduction in flowering plants involves two fertilization events in the haploid female gametophyte. One sperm nucleus fertilizes the egg to form the embryo. A second sperm nucleus fertilizes the central cell to form the endosperm, a unique tissue that supports the growth of the embryo. Fertilization also activates maternal tissue differentiation, the ovule integuments form the seed coat and the ovary forms the fruit.

10       The present invention is based, at least in part, on the discovery of a set of female-gametophytic mutations and the subsequent cloning of the gene involved, termed *DEMETER* (*DMT*), formally known as *ATROPOS* (*ATR*). Two mutant alleles of *DMT* disclosed here were created using a T-DNA tag, thereby disrupting an exon of the gene. The *dmt* mutations affect endosperm production, allowing for increased endosperm  
15       development. Generally, the mutant *dmt* alleles are not transmitted by the female gametophyte. Inheritance of a mutant *dmt* allele by the female gametophyte usually results in embryo abortion and endosperm overproduction, even when the pollen bears the wild-type *DMT* allele.

20       In contrast, transmission of *dmt* mutant alleles through the male gametophyte (*i.e.*, pollen) is ecotype-dependent in *Arabidopsis*. For instance, in some ecotypes (e.g., Columbia), transmission of *dmt* mutant alleles is less than 50%. However, in Landsberg *erecta*, transmission is almost normal.

25       DMT is a repressor of endosperm both before and after fertilization. DMT is both necessary and sufficient for MEDEA transcription. DMT is related to 5-methylcytosine glycosylases. DMT regulates transcription of specific target genes (*i.e.*, MEA) by a demethylation mechanism. DMT is also required for maintaining the proper global pattern of methylation of chromosomal DNA in cells.

30       The isolated sequences prepared as described herein, can be used in a number of techniques, for example, to suppress or enhance endogenous DMT gene expression. Modulation of *DMT* gene expression or DMT activity in plants is particularly useful, for example, in producing embryo-less or embryo-reduced seed, seed with increased endosperm, as part of a system to generate seed, to modulate time to flowering, organ identity, size and/or number, meristem size or activity in plants, or to modulate methylation, and thus gene expression in plants. Another use is the expression of DMT

polynucleotides in animal cells, for instance as a DNA repair enzyme useful in preventing the unnatural proliferation of cells (including cancer) due to chromosomal lesions. See, e.g., Bruner, *et al.*, *Nature* 403:859 (2000).

As described in more detail below, reduction of expression of DMT in plants results in a number of diverse phenotypes. Without intending to limit the invention to particular embodiments, it is believed that some of the phenotypes that are generated in plants are epigenetic mutations, i.e., effects due to differences in the methylation state of the chromosome that result in altered gene expression. Thus, DMT provides a powerful tool to develop any number of plant lines with a variety of desired phenotypes.

### ***Isolation of DMT nucleic acids***

Generally, the nomenclature and the laboratory procedures in recombinant DNA technology described below are those well known and commonly employed in the art. Standard techniques are used for cloning, DNA and RNA isolation, amplification and purification. Generally enzymatic reactions involving DNA ligase, DNA polymerase, restriction endonucleases and the like are performed according to the manufacturer's specifications. These techniques and various other techniques are generally performed according to Sambrook *et al.*, *Molecular Cloning - A Laboratory Manual*, Cold Spring Harbor Laboratory, Cold Spring Harbor, New York, (1989).

The isolation of *DMT* nucleic acids may be accomplished by a number of techniques. For instance, oligonucleotide probes based on the sequences disclosed here can be used to identify the desired gene in a cDNA or genomic DNA library. To construct genomic libraries, large segments of genomic DNA are generated by random fragmentation, e.g. using restriction endonucleases, and are ligated with vector DNA to form concatemers that can be packaged into the appropriate vector. To prepare a cDNA library, mRNA is isolated from the desired organ, such as ovules, and a cDNA library which contains the *DMT* gene transcript is prepared from the mRNA. Alternatively, cDNA may be prepared from mRNA extracted from other tissues in which *DMT* genes or homologs are expressed.

The cDNA or genomic library can then be screened using a probe based upon the sequence of a cloned *DMT* gene disclosed here. Probes may be used to hybridize with genomic DNA or cDNA sequences to isolate homologous genes in the same or different plant species. Alternatively, antibodies raised against a DMT polypeptide can be used to screen an mRNA expression library.

Alternatively, the nucleic acids of interest can be amplified from nucleic acid samples using amplification techniques. For instance, polymerase chain reaction (PCR) technology can be used to amplify the sequences of the *DMT* genes directly from genomic DNA, from cDNA, from genomic libraries or cDNA libraries. PCR and other *in vitro* amplification methods may also be useful, for example, to clone nucleic acid sequences that code for proteins to be expressed, to make nucleic acids to use as probes for detecting the presence of the desired mRNA in samples, for nucleic acid sequencing, or for other purposes. For a general overview of PCR see *PCR Protocols: A Guide to Methods and Applications*. (Innis, M, Gelfand, D., Sninsky, J. and White, T., eds.), Academic Press, San Diego (1990).

Appropriate primers and probes for identifying *DMT* sequences from plant tissues are generated from comparisons of the sequences provided here with other related genes. For instance, *DMT* can be compared to the other endonuclease III genes, such as Genbank Accession No. AE002073. Using these techniques, one of skill can identify conserved regions in the nucleic acids disclosed here to prepare the appropriate primer and probe sequences. Primers that specifically hybridize to conserved regions in *DMT* genes can be used to amplify sequences from widely divergent plant species. Appropriate primers for amplification of the genomic region or cDNA of *DMT* include the following primers:

Xba-SKEN-7; CCTCTAGAGGAATTGTCGGCAAAATCGAG  
 SKB-8; GGAGAGACGGTTATTGTCAACC  
 SKB-7; AAAAGTCTACAAGGGAGAGAGAGT  
 SKB-5; GTAGATGTACATACGTACC  
 SKEN-8; GCATCCTCCAACAAGTAACAATCCACTC  
 SKB-6; CACTGAGATTAATTCTTCAGACTCG  
 SKEN-3.5; CTCAGGCGAGTCAATGCCGGAGAACAC  
 SKEN-3; CGAGGGCTGATCCGGGGATAGATATTTT  
 SKEN-2; CCCCCGATCAGCCCTCGAATTC  
 SKEN-1; CCCCTGTCTACAAATTCACCACCTGG  
 SKEL-4; CTGACCCAACTGCTTCTCTTC  
 skes1.5; TCACCTGTTCTGAACAGACTGG  
 SKES-1.4; CAGCAGACGAGTCCATAATGCTCTGC  
 SKES-2.4; GGTTCGCTTCCACGACCACC  
 SKES-1; GGAAGCCACGCAAAGCTGCAACTCAGG  
 SKES-2.45; GAGTTGCAGCTTTGCGTGGCTTCC



SKES2.5; TTCAGACTCAGAGTCACCTTGC  
 SKES-2; ACCAGCAGCCTTGCTTGGCC  
 SKES-3; CATGCCAGAGAAGCAGGGCTCC  
 SKES3.5; CGATGATACTGTCTCTTCGAGC  
 SKES-6; CCTCCGCCTGCTCATGCCTCAG  
 SKEN-4; GTCCATCAGGAGAACTTCTGTGCAGGAT  
 SKES-4; GGGAAACAAGTGCACCATCTCC  
 SKEN-6; GCTCTCATAGGGAACAAGTGCACCATCTC  
 SKES-5; CGCTCGCATGCACCTGGTAC  
 SKB-1; GGAGGGAATCGAGCAGCTAGAG  
 SKB-2; GAGCAGCTAAGGGACTGTTCAAACCT  
 SKB-3; CCAGGAATGGGATTGTCCGG  
 3' RACE-2; CTTGGACGCGCTTGAGGAACC  
 3' RACE-1; GCCTACAAGCCAGTGGGATAG  
 cDNA-1; GCCAAGGACTATCTCTTGAGC  
 SKB-4; GGATGGACTCGAGCACTGGG  
 SKE2.2-4; AGAGGAGAGTGCAGACACTTTG  
 cDNA-3; GAGGACCCTGACGAGATCCCAAC  
 cDNA-9; CCATGTGTTCCCGTAGAGTCATTCC  
 2.2+SKE-1; ATGGAGCTCCAAGAAGGTGACATG  
 cDNA-5; CAGAAGTGTGGAGGGAAAGCGTCTGGC  
 cDNA-4; CCCTCAGACTGTTACACTCAGAAC  
 cDNA-2; CCCGTGAGCGGAAAACCTCTCATGGC  
 cDNA-7; GGAAAGGATTTCGTATGTGTCCGTGG  
 SKEN-5; GCAATGCGTTTGCTTTCTTCCAGTCATCT  
 cDNA-6; GAGGAGAGCAGAGAAGCAATGCGTTTGC  
 cDNA-8; GTTAGAGAGAAAATAAATAACCC  
 2.2+SKE-3; CCGTAAACAACACCGGATACAC

The amplification conditions are typically as follows. Reaction components: 10 mM Tris-HCl, pH 8.3, 50 mM potassium chloride, 1.5 mM magnesium chloride, 0.001% gelatin, 200  $\mu$ M dATP, 200  $\mu$ M dCTP, 200  $\mu$ M dGTP, 200  $\mu$ M dTTP, 0.4  $\mu$ M primers, and 100 units per ml Taq polymerase. Program: 96 C for 3 min., 30 cycles of 96 C for 45 sec., 50 C for 60 sec., 72 for 60 sec, followed by 72 C for 5 min.

Standard nucleic acid hybridization techniques using the conditions disclosed above can then be used to identify full-length cDNA or genomic clones.

Alternatively, a number of methods for designing modifications of polynucleotide sequences are known to those of skill in the art. For example, oligonucleotide directed mutagenesis can be used to introduce site-specific mutations in a nucleic acid sequence of interest. Examples of such techniques are found in the references above and, e.g., in Reidhaar-Olson *et al. Science*, 241:53-57 (1988) and Ausubel *et al.* Similarly, gene shuffling (Stemmer *Proc. Natl. Acad. Sci. USA* 91:10747-10751(1994); Ostermeier *et al. Proc. Natl. Acad. Sci. USA*, 96: 3562-67(1999))) can be used to introduce variation into one or more *DMT* sequences or subsequences. For example, orthologous (between species) or homologous (within a species) *DMT* nucleic acids can be interchanged, combined or shuffled to produce novel variations within the scope of the invention.

Additionally, error prone PCR can also be used to introduce variation into a nucleic acid sequence. See, Leung *et al.* (1989) *Technique* 1:11-15 and Caldwell *et al.* (1992) *PCR Methods Applic.* 2:28-33.

#### ***Control of DMT activity or gene expression***

Since *DMT* genes are involved in controlling seed, in particular endosperm, development, inhibition of endogenous *DMT* activity or gene expression is useful in a number of contexts. For instance, reduction of *DMT* activity can be used for production of seed with enhanced endosperm. By reducing and/or eliminating *DMT* activity, plants with seed containing increased endosperm can be produced.

Alternatively, substantial inhibition of *DMT* activity can be used for production of fruit with small and/or degraded seed (referred to here as "seedless fruit") after fertilization. In many plants, particularly dicots, the endosperm is not persistent and eventually is degraded. Thus, in plants of the invention in which *DMT* activity is inhibited, embryo-less seed do not persist and seedless fruit are produced. For production of dicots with enhanced endosperm, the most beneficial effect may be to reduce, but not eliminate *DMT* activity. On the other hand, in monocots, which have persistent endosperm, it is advantageous to eliminate *DMT* activity.

Alternatively, plants of the invention can be used to prevent pre-harvest sprouting in seeds, especially those derived from cereals. In these plants, the endosperm persists and is the major component of the mature seed. Premature growth of embryos in stored grain causes release of degradative enzymes which digest starch and other

components of the endosperm. Plants of the present invention are useful in addressing this problem because the seeds lack an embryo and thus will not germinate.

Moreover, as discussed herein, time to flowering and DNA methylation can also be modulated by modulating DMT activity in a cell. For example, DMT can be used to modulate the amount of methylated DNA in a cell. Indeed, since expression of many genes is dependent on their methylation state, modulation of DMT activity modulates gene expression in a cell. Examples of genes whose expression is modulated by DMT include *MEDEA*.

One of skill will recognize that a number of methods can be used to modulate DMT activity or gene expression. DMT activity can be modulated in the plant cell at the gene, transcriptional, posttranscriptional, translational, or posttranslational, levels. Techniques for modulating DMT activity at each of these levels are generally well known to one of skill and are discussed briefly below.

Methods for introducing genetic mutations into plant genes are well known. For instance, seeds or other plant material can be treated with a mutagenic chemical substance, according to standard techniques. Such chemical substances include, but are not limited to, the following: diethyl sulfate, ethylene imine, ethyl methanesulfonate and N-nitroso-N-ethylurea. Alternatively, ionizing radiation from sources such as, for example, X-rays or gamma rays can be used.

Alternatively, homologous recombination can be used to induce targeted gene disruptions by specifically deleting or altering the *DMT* gene *in vivo* (see, generally, Grewal and Klar, *Genetics* 146: 1221-1238 (1997) and Xu *et al.*, *Genes Dev.* 10:2411-2422 (1996)). Homologous recombination has been demonstrated in plants (Puchta *et al.*, *Experientia* 50:277-284 (1994), Swoboda *et al.*, *EMBO J.* 13:484-489 (1994); Offringa *et al.*, *Proc. Natl. Acad. Sci. USA* 90: 7346-7350 (1993); and Kempin *et al.* *Nature* 389:802-803 (1997)).

In applying homologous recombination technology to the genes of the invention, mutations in selected portions of a *DMT* gene sequences (including 5' upstream, 3' downstream, and intragenic regions) such as those disclosed here are made *in vitro* and then introduced into the desired plant using standard techniques. Since the efficiency of homologous recombination is known to be dependent on the vectors used, use of dicistronic gene targeting vectors as described by Mountford *et al.* *Proc. Natl. Acad. Sci. USA* 91:4303-4307 (1994); and Vaulont *et al.* *Transgenic Res.* 4:247-255 (1995) are conveniently used to increase the efficiency of selecting for altered *DMT* gene

expression in transgenic plants. The mutated gene will interact with the target wild-type gene in such a way that homologous recombination and targeted replacement of the wild-type gene will occur in transgenic plant cells, resulting in suppression of DMT activity.

Alternatively, oligonucleotides composed of a contiguous stretch of RNA and DNA residues in a duplex conformation with double hairpin caps on the ends can be used. The RNA/DNA sequence is designed to align with the sequence of the target *DMT* gene and to contain the desired nucleotide change. Introduction of the chimeric oligonucleotide on an extrachromosomal T-DNA plasmid results in efficient and specific DMT gene conversion directed by chimeric molecules in a small number of transformed plant cells. This method is described in Cole-Strauss *et al. Science* 273:1386-1389 (1996) and Yoon *et al. Proc. Natl. Acad. Sci. USA* 93:2071-2076 (1996).

Gene expression can be inactivated using recombinant DNA techniques by transforming plant cells with constructs comprising transposons or T-DNA sequences. *DMT* mutants prepared by these methods are identified according to standard techniques. For instance, mutants can be detected by PCR or by detecting the presence or absence of *DMT* mRNA, *e.g.*, by Northern blots. Mutants can also be selected by assaying for development of endosperm in the absence of fertilization.

The isolated nucleic acid sequences prepared as described herein, can also be used in a number of techniques to control endogenous *DMT* gene expression at various levels. Subsequences from the sequences disclosed here can be used to control, transcription, RNA accumulation, translation, and the like.

A number of methods can be used to inhibit gene expression in plants. For instance, antisense technology can be conveniently used. To accomplish this, a nucleic acid segment from the desired gene is cloned and operably linked to a promoter such that the antisense strand of RNA will be transcribed. The construct is then transformed into plants and the antisense strand of RNA is produced. In plant cells, it has been suggested that antisense suppression can act at all levels of gene regulation including suppression of RNA translation (*see, Bourque Plant Sci. (Limerick)* 105:125-149 (1995); Pantopoulos In Progress in Nucleic Acid Research and Molecular Biology, Vol. 48. Cohn, W. E. and K. Moldave (Ed.). Academic Press, Inc.: San Diego, California, USA; London, England, UK. p. 181-238; Heiser *et al. Plant Sci. (Shannon)* 127:61-69 (1997)) and by preventing the accumulation of mRNA which encodes the protein of interest, (*see, Baulcombe Plant Mol. Bio.* 32:79-88 (1996); Prins and Goldbach *Arch. Virol.* 141:2259-2276 (1996);

Metzlaff *et al. Cell* 88:845-854 (1997), Sheehy *et al., Proc. Nat. Acad. Sci. USA*, 85:8805-8809 (1988), and Hiatt *et al., U.S. Patent No. 4,801,340*.

The nucleic acid segment to be introduced generally will be substantially identical to at least a portion of the endogenous *DMT* gene or genes to be repressed. The sequence, however, need not be perfectly identical to inhibit expression. The vectors of the present invention can be designed such that the inhibitory effect applies to other genes within a family of genes exhibiting homology or substantial homology to the target gene.

For antisense suppression, the introduced sequence also need not be full length relative to either the primary transcription product or fully processed mRNA. Generally, higher homology can be used to compensate for the use of a shorter sequence. Furthermore, the introduced sequence need not have the same intron or exon pattern, and homology of non-coding segments may be equally effective. Normally, a sequence of between about 30 or 40 nucleotides and about full length nucleotides should be used, though a sequence of at least about 100 nucleotides is preferred, a sequence of at least about 200 nucleotides is more preferred, and a sequence of about 500 to about 7000 nucleotides is especially preferred.

A number of gene regions can be targeted to suppress *DMT* gene expression. The targets can include, for instance, the coding regions, introns, sequences from exon/intron junctions, 5' or 3' untranslated regions, and the like. In some embodiments, the constructs can be designed to eliminate the ability of regulatory proteins to bind to *DMT* gene sequences that are required for its cell- and/or tissue-specific expression. Such transcriptional regulatory sequences can be located either 5'-, 3'-, or within the coding region of the gene and can be either promote (positive regulatory element) or repress (negative regulatory element) gene transcription. These sequences can be identified using standard deletion analysis, well known to those of skill in the art. Once the sequences are identified, an antisense construct targeting these sequences is introduced into plants to control gene transcription in particular tissue, for instance, in developing ovules and/or seed. In one embodiment, transgenic plants are selected for *DMT* activity that is reduced but not eliminated.

Oligonucleotide-based triple-helix formation can be used to disrupt *DMT* gene expression. Triplex DNA can inhibit DNA transcription and replication, generate site-specific mutations, cleave DNA, and induce homologous recombination (*see, e.g.,* Havre and Glazer *J. Virology* 67:7324-7331 (1993); Scanlon *et al. FASEB J.* 9:1288-1296 (1995); Giovannangeli *et al. Biochemistry* 35:10539-10548 (1996); Chan and Glazer *J.*

*Mol. Medicine (Berlin)* 75:267-282 (1997)). Triple helix DNAs can be used to target the same sequences identified for antisense regulation.

Catalytic RNA molecules or ribozymes can also be used to inhibit expression of *DMT* genes. It is possible to design ribozymes that specifically pair with virtually any target RNA and cleave the phosphodiester backbone at a specific location, thereby functionally inactivating the target RNA. In carrying out this cleavage, the ribozyme is not itself altered, and is thus capable of recycling and cleaving other molecules, making it a true enzyme. The inclusion of ribozyme sequences within antisense RNAs confers RNA-cleaving activity upon them, thereby increasing the activity of the constructs. Thus, ribozymes can be used to target the same sequences identified for antisense regulation.

A number of classes of ribozymes have been identified. One class of ribozymes is derived from a number of small circular RNAs which are capable of self-cleavage and replication in plants. The RNAs replicate either alone (viroid RNAs) or with a helper virus (satellite RNAs). Examples include RNAs from avocado sunblotch viroid and the satellite RNAs from tobacco ringspot virus, lucerne transient streak virus, velvet tobacco mottle virus, solanum nodiflorum mottle virus and subterranean clover mottle virus. The design and use of target RNA-specific ribozymes is described in Zhao and Pick *Nature* 365:448-451 (1993); Eastham and Ahlering *J. Urology* 156:1186-1188 (1996); Sokol and Murray *Transgenic Res.* 5:363-371 (1996); Sun *et al. Mol. Biotechnology* 7:241-251 (1997); and Haseloff *et al. Nature*, 334:585-591 (1988).

Another method of suppression is sense cosuppression. Introduction of nucleic acid configured in the sense orientation has been recently shown to be an effective means by which to block the transcription of target genes. For an example of the use of this method to modulate expression of endogenous genes (see, Assaad *et al. Plant Mol. Bio.* 22:1067-1085 (1993); Flavell *Proc. Natl. Acad. Sci. USA* 91:3490-3496 (1994); Stam *et al. Annals Bot.* 79:3-12 (1997); Napoli *et al., The Plant Cell* 2:279-289 (1990); and U.S. Patents Nos. 5,034,323, 5,231,020, and 5,283,184).

The suppressive effect may occur where the introduced sequence contains no coding sequence *per se*, but only intron or untranslated sequences homologous to sequences present in the primary transcript of the endogenous sequence. The introduced sequence generally will be substantially identical to the endogenous sequence intended to be repressed. This minimal identity will typically be greater than about 65%, but a higher identity might exert a more effective repression of expression of the endogenous

sequences. Substantially greater identity of more than about 80% is preferred, though about 95% to absolute identity would be most preferred. As with antisense regulation, the effect should apply to any other proteins within a similar family of genes exhibiting homology or substantial homology.

5 For sense suppression, the introduced sequence, needing less than absolute identity, also need not be full length, relative to either the primary transcription product or fully processed mRNA. This may be preferred to avoid concurrent production of some plants that are overexpressers. A higher identity in a shorter than full length sequence compensates for a longer, less identical sequence. Furthermore, the introduced sequence  
10 need not have the same intron or exon pattern, and identity of non-coding segments will be equally effective. Normally, a sequence of the size ranges noted above for antisense regulation is used. In addition, the same gene regions noted for antisense regulation can be targeted using cosuppression technologies.

In a preferred embodiment, expression of a nucleic acid of interest can be  
15 suppressed by the simultaneous expression of both sense and antisense constructs (Waterhouse *et al.*, *Proc. Natl. Acad. Sci. USA* 95:13959-13964 (1998). See also Tabara *et al. Science* 282:430-431 (1998).

Alternatively, *DMT* activity may be modulated by eliminating the proteins that are required for *DMT* cell-specific gene expression. Thus, expression of regulatory  
20 proteins and/or the sequences that control *DMT* gene expression can be modulated using the methods described here.

Another method is use of engineered tRNA suppression of *DMT* mRNA translation. This method involves the use of suppressor tRNAs to transactivate target genes containing premature stop codons (see, Betzner *et al. Plant J* 11:587-595 (1997);  
25 and Choise *et al. Plant J* 11:597-604 (1997). A plant line containing a constitutively expressed *DMT* gene that contains an amber stop codon is first created. Multiple lines of plants, each containing tRNA suppressor gene constructs under the direction of cell-type specific promoters are also generated. The tRNA gene construct is then crossed into the *DMT* line to activate *DMT* activity in a targeted manner. These tRNA suppressor lines  
30 could also be used to target the expression of any type of gene to the same cell or tissue types.

*DMT* proteins may form homogeneous or heterologous complexes *in vivo*. Thus, production of dominant-negative forms of *DMT* polypeptides that are defective in their abilities to bind to other proteins in the complex is a convenient means to inhibit

09840743-042301

endogenous DMT activity. This approach involves transformation of plants with constructs encoding mutant DMT polypeptides that form defective complexes and thereby prevent the complex from forming properly. The mutant polypeptide may vary from the naturally occurring sequence at the primary structure level by amino acid substitutions, additions, deletions, and the like. These modifications can be used in a number of combinations to produce the final modified protein chain. Use of dominant negative mutants to inactivate target genes is described in Mizukami *et al. Plant Cell* 8:831-845 (1996).

Another strategy to affect the ability of a DMT protein to interact with itself or with other proteins involves the use of antibodies specific to DMT. In this method cell-specific expression of DMT-specific Abs is used inactivate functional domains through antibody:antigen recognition (*see*, Hupp *et al. Cell* 83:237-245 (1995)).

After plants with reduced DMT activity are identified, a recombinant construct capable of expressing low levels of DMT in embryos can be introduced using the methods discussed below. In this fashion, the level of DMT activity can be regulated to produce preferred plant phenotypes. For example, a relatively weak promoter such as the ubiquitin promoter (*see*, e.g., Garbarino *et al. Plant Physiol.* 109(4):1371-8 (1995); Christensen *et al Transgenic Res.* 5(3):213-8 (1996); and Holtorf *et al. Plant. Mol. Biol.* 29(4):637-46 (1995)) is useful to produce plants with reduced levels of DMT activity or expression. Such plants are useful for producing, for instance, plants that produce seed with enhanced endosperm.

### ***Use of nucleic acids of the invention to enhance DMT gene expression***

Isolated sequences prepared as described herein can also be introduced into a plant cell, thereby modulating expression of a particular *DMT* nucleic acid to enhance or increase endogenous gene expression. For instance, without being bound to any theory, in light of DMT's relation to Exonuclease III and DNA glycosylases, applicants believe that DMT binds DNA or chromatin and acts to modulate transcription by modulating the methylation state of DNA. Enhanced expression can therefore be used to control plant morphology by controlling expression of genes under DMT's control, such as *MEDEA*, in desired tissues or cells. Enhanced expression can also be used, for instance, to increase vegetative growth by preventing the plant from setting seed. Where overexpression of a gene is desired, the desired gene from a different species may be used to decrease potential sense suppression effects.



Moreover, as discussed herein, time to flowering and DNA methylation can also be modulated by modulating DMT activity in a cell. For example, increased expression of DMT in a plant results in delayed time to flowering. Similarly, DMT can be used to modulate the amount of methylated DNA in a cell. Indeed, since expression of many genes is dependent on their methylation state, modulation of DMT activity modulates gene expression in a cell. Examples of genes whose expression is modulated by DMT include *MEDEA*.

One of skill will recognize that the polypeptides encoded by the genes of the invention, like other proteins, have different domains that perform different functions. Thus, the gene sequences need not be full length, so long as the desired functional domain of the protein is expressed.

Modified protein chains can also be readily designed utilizing various recombinant DNA techniques well known to those skilled in the art and described in detail, below. For example, the chains can vary from the naturally occurring sequence at the primary structure level by amino acid substitutions, additions, deletions, and the like. These modifications can be used in a number of combinations to produce the final modified protein chain.

#### ***Preparation of recombinant vectors***

To use isolated sequences in the above techniques, recombinant DNA vectors suitable for transformation of plant cells are prepared. Techniques for transforming a wide variety of flowering plant species are well known and described in the technical and scientific literature. See, for example, Weising *et al. Ann. Rev. Genet.* 22:421-477 (1988). A DNA sequence coding for the desired polypeptide, for example a cDNA sequence encoding a full length protein, will preferably be combined with transcriptional and translational initiation regulatory sequences which will direct the transcription of the sequence from the gene in the intended tissues of the transformed plant.

For example, for overexpression, a plant promoter fragment may be employed which will direct expression of the gene in all tissues of a regenerated plant. Such promoters are referred to herein as "constitutive" promoters and are active under most environmental conditions and states of development or cell differentiation. Examples of constitutive promoters include the cauliflower mosaic virus (CaMV) 35S transcription initiation region, the 1'- or 2'- promoter derived from T-DNA of

*Agrobacterium tumefaciens*, and other transcription initiation regions from various plant genes known to those of skill. Such genes include for example, *ACT11* from *Arabidopsis* (Huang *et al. Plant Mol. Biol.* 33:125-139 (1996)), *Cat3* from *Arabidopsis* (GenBank No. U43147, Zhong *et al., Mol. Gen. Genet.* 251:196-203 (1996)), the gene encoding

5 stearoyl-acyl carrier protein desaturase from *Brassica napus* (Genbank No. X74782, Solocombe *et al. Plant Physiol.* 104:1167-1176 (1994)), *Gpc1* from maize (GenBank No. X15596, Martinez *et al. J. Mol. Biol.* 208:551-565 (1989)), and *Gpc2* from maize (GenBank No. U45855, Manjunath *et al., Plant Mol. Biol.* 33:97-112 (1997)).

Alternatively, the plant promoter may direct expression of the DMT

10 nucleic acid in a specific tissue or may be otherwise under more precise environmental or developmental control. Examples of environmental conditions that may effect transcription by inducible promoters include anaerobic conditions, elevated temperature, or the presence of light. Such promoters are referred to here as "inducible" or "tissue-specific" promoters. One of skill will recognize that a tissue-specific promoter may drive

15 expression of operably linked sequences in tissues other than the target tissue. Thus, as used herein a tissue-specific promoter is one that drives expression preferentially in the target tissue, but may also lead to some expression in other tissues as well.

Examples of promoters under developmental control include promoters that initiate transcription only (or primarily only) in certain tissues, such as fruit, seeds, or

20 flowers. Promoters that direct expression of nucleic acids in ovules, flowers or seeds are particularly useful in the present invention. As used herein a seed-specific promoter is one which directs expression in seed tissues, such promoters may be, for example, ovule-specific (which includes promoters which direct expression in maternal tissues or the female gametophyte, such as egg cells or the central cell), embryo-specific, endosperm-specific, integument-specific, seed coat-specific, or some combination thereof. Examples

25 include a promoter from the ovule-specific *BEL1* gene described in Reiser *et al. Cell* 83:735-742 (1995) (GenBank No. U39944). Other suitable seed specific promoters are derived from the following genes: *MAC1* from maize (Sheridan *et al. Genetics* 142:1009-1020 (1996)), *Cat3* from maize (GenBank No. L05934, Abler *et al. Plant Mol. Biol.*

30 22:10131-1038 (1993), the gene encoding oleosin 18kD from maize (GenBank No. J05212, Lee *et al. Plant Mol. Biol.* 26:1981-1987 (1994)), viviparous-1 from *Arabidopsis* (Genbank No. U93215), the gene encoding oleosin from *Arabidopsis* (Genbank No. Z17657), *Atmyc1* from *Arabidopsis* (Urao *et al. Plant Mol. Biol.* 32:571-576 (1996)), the 2s seed storage protein gene family from *Arabidopsis* (Conceicao *et al. Plant* 5:493-505

09840743-042301  
TOP SECRET FRODO BAGGINS

(1994)) the gene encoding oleosin 20kD from *Brassica napus* (GenBank No. M63985), *napA* from *Brassica napus* (GenBank No. J02798, Josefsson *et al.* *JBL* 26:12196-1301 (1987), the napin gene family from *Brassica napus* (Sjodahl *et al.* *Planta* 197:264-271 (1995), the gene encoding the 2S storage protein from *Brassica napus* (Dasgupta *et al.* *Gene* 133:301-302 (1993)), the genes encoding oleosin A (Genbank No. U09118) and oleosin B (Genbank No. U09119) from soybean and the gene encoding low molecular weight sulphur rich protein from soybean (Choi *et al.* *Mol Gen. Genet.* 246:266-268 (1995)).

In addition, the promoter sequences from the *DMT* genes disclosed here can be used to drive expression of the *DMT* polynucleotides of the invention or heterologous sequences. The sequences of the promoters are identified below.

If proper polypeptide expression is desired, a polyadenylation region at the 3'-end of the coding region should be included. The polyadenylation region can be derived from the natural gene, from a variety of other plant genes, or from T-DNA.

The vector comprising the sequences (*e.g.*, promoters or coding regions) from genes of the invention will typically comprise a marker gene which confers a selectable phenotype on plant cells. For example, the marker may encode biocide resistance, particularly antibiotic resistance, such as resistance to kanamycin, G418, bleomycin, hygromycin, or herbicide resistance, such as resistance to chlorosulfuron or Basta.

### ***Promoter and Enhancer Nucleic Acids of the Invention***

The present invention provides polynucleotides useful as promoters and enhancers. The invention also provides methods of targeting heterologous polypeptides to a female gametophyte of a plant, including, *e.g.*, the polar nuclei, the eggs and synergids and central cells. Promoter polynucleotides of the invention include, for example, sequences and subsequences of the *DMT* 5' flanking DNA (SEQ ID NO:3), the 5' UTR region (SEQ ID NO:6) and the 3' flanking region (SEQ ID NO:4). In some embodiments, the promoter sequences are operably linked to the 5' end of the *DMT* coding region, which is in turn fused to a polynucleotide of interest, typically encoding a polypeptide. An exemplary promoter sequence includes the last 3424 nucleotides of SEQ ID NO:3 linked to the first 1478 nucleotides of SEQ ID NO:5. In some embodiments, a further 444 nucleotides (*e.g.*, the first 444 nucleotides of the *DMT* coding region) are incorporated into the promoter. In some embodiments, the promoter

sequences of the invention specifically direct expression of polynucleotides to the female gametophyte and does not direct expression in tissues following fertilization.

### ***Production of transgenic plants***

5 DNA constructs of the invention may be introduced into the genome of the desired plant host by a variety of conventional techniques. For example, the DNA construct may be introduced directly into the genomic DNA of the plant cell using techniques such as electroporation and microinjection of plant cell protoplasts, or the DNA constructs can be introduced directly to plant tissue using ballistic methods, such as  
10 DNA particle bombardment.

Microinjection techniques are known in the art and well described in the scientific and patent literature. The introduction of DNA constructs using polyethylene glycol precipitation is described in Paszkowski et al. *Embo J.* 3:2717-2722 (1984). Electroporation techniques are described in Fromm et al. *Proc. Natl. Acad. Sci. USA*  
15 82:5824 (1985). Ballistic transformation techniques are described in Klein et al. *Nature* 327:70-73 (1987).

Alternatively, the DNA constructs may be combined with suitable T-DNA flanking regions and introduced into a conventional *Agrobacterium tumefaciens* host vector. The virulence functions of the *Agrobacterium tumefaciens* host will direct the  
20 insertion of the construct and adjacent marker into the plant cell DNA when the cell is infected by the bacteria. *Agrobacterium tumefaciens*-mediated transformation techniques, including disarming and use of binary vectors, are well described in the scientific literature. See, for example Horsch et al. *Science* 233:496-498 (1984), and Fraley et al. *Proc. Natl. Acad. Sci. USA* 80:4803 (1983).

25 Transformed plant cells which are derived by any of the above transformation techniques can be cultured to regenerate a whole plant which possesses the transformed genotype and thus the desired phenotype such as increased seed mass. Such regeneration techniques rely on manipulation of certain phytohormones in a tissue culture growth medium, typically relying on a biocide and/or herbicide marker which has been  
30 introduced together with the desired nucleotide sequences. Plant regeneration from cultured protoplasts is described in Evans et al., *Protoplasts Isolation and Culture, Handbook of Plant Cell Culture*, pp. 124-176, MacMillan Publishing Company, New York, 1983; and Binding, *Regeneration of Plants, Plant Protoplasts*, pp. 21-73, CRC Press, Boca Raton, 1985. Regeneration can also be obtained from plant callus, explants,

organs, or parts thereof. Such regeneration techniques are described generally in Klee et al. *Ann. Rev. of Plant Phys.* 38:467-486 (1987).

The nucleic acids of the invention can be used to confer desired traits on essentially any plant. Thus, the invention has use over a broad range of plants, including

5 species from the genera *Anacardium*, *Arachis*, *Asparagus*, *Atropa*, *Avena*, *Brassica*, *Citrus*, *Citrullus*, *Capsicum*, *Carthamus*, *Cocos*, *Coffea*, *Cucumis*, *Cucurbita*, *Daucus*, *Elaeis*, *Fragaria*, *Glycine*, *Gossypium*, *Helianthus*, *Heterocallis*, *Hordeum*, *Hyoscyamus*, *Lactuca*, *Linum*, *Lolium*, *Lupinus*, *Lycopersicon*, *Malus*, *Manihot*, *Majorana*, *Medicago*, *Nicotiana*, *Olea*, *Oryza*, *Panicum*, *Pannisetum*, *Persea*, *Phaseolus*, *Pistachia*, *Pisum*,

10 *Pyrus*, *Prunus*, *Raphanus*, *Ricinus*, *Secale*, *Senecio*, *Sinapis*, *Solanum*, *Sorghum*, *Theobromus*, *Trigonella*, *Triticum*, *Vicia*, *Vitis*, *Vigna*, and *Zea*.

One of skill will recognize that after the expression cassette is stably incorporated in transgenic plants and confirmed to be operable, it can be introduced into other plants by sexual crossing. Any of a number of standard breeding techniques can be

15 used, depending upon the species to be crossed.

Seed obtained from plants of the present invention can be analyzed according to well known procedures to identify plants with the desired trait. If antisense or other techniques are used to control DMT gene expression, Northern blot analysis can be used to screen for desired plants. In addition, the presence of fertilization independent

20 reproductive development can be detected. Plants can be screened, for instance, for the ability to form embryo-less seed, form seed that abort after fertilization, or set fruit in the absence of fertilization. These procedures will depend, part on the particular plant species being used, but will be carried out according to methods well known to those of skill.

### **DMT Mutations, Fragments And Fusions**

As discussed above, DMT polynucleotides and polypeptides are not limited to the sequences disclosed herein. Those of skill in the art that conservative amino acid substitutions, as well as amino acid additions or deletions may not result in

30 any change in biological activity. Moreover, sequence variants with at least one modulated biological activity of DMT are also contemplated. For example, at least one DMT activity can be increased or decreased by introduction of single or multiple amino acid changes from the sequences disclosed herein. Those of skill in the art will recognize that conservative amino acid substitutions in important functional domains are typically

useful in generating more active DMT polypeptides. Conversely, non-conservative substitutions of amino acid residues in functional domains, such as the HhH region of DMT (e.g., amino acids 1271-1304 of SEQ ID NO:2) are likely to disrupt at least one biological activity such as DNA binding. In some embodiments, the fragments of the invention consist of about 100, 200, 300 400, 500, 600, 700, 800, 900, or 1000 amino acids.

Alternatively, fragments of the sequences disclosed herein are contemplated. In some preferred embodiments, the polypeptide fragments have at least one biological activity of DMT. For example, amino acid sequences comprising DMT domain B represent polypeptide fragments with glycosylase or demethylase activity. In some embodiments, a fragment comprising amino acids 1167-1404, 1192-1404, 1192-1368 or 1167-1368 of SEQ ID NO:2 have glycosylase activity.

Mutations, fragments and fusions are also useful as dominant negative mutations. For instance, different regions of the DMT protein are responsible for different biological activities. Thus, mutation or deletion of one functional domain can eliminate one but not all activities. For example, mutation or deletion of the DNA binding domain may result in proteins that interact with proteins necessary for DMT function, effectively titrating out those proteins and preventing an active DMT protein from acting. Similarly, DMT fragments comprising the DNA binding portion of the protein with an inactive enzymatic domain or lacking an enzymatic domain are also useful as dominant negative mutants by competing with active DMT polypeptides for DNA binding sites. As described herein, domains of DMT that can be modulated include: the leucine zipper, nuclear localization sequence, HhH domain, the aspartic acid of the GPD domain, as well as DMT domains A, B or C. Without intending to limit the scope of the invention, based on the data provided herein, DMT has glycosylase and demethylase activity and is a DNA repair enzyme.

#### ***Targeting the polypeptides of the invention to chromosomal regions***

Without intending to limit the scope of the invention, based on the data provided herein, it is believed that DMT has glycosylase and/or demethylase activity and is a DNA repair enzyme. DNA methylation plays an important role in the repression of gene transcription during animal development including embryogenesis, myogenesis and blood cell development. Methylated DNA is recognized by MeCP2 which in turn represses gene transcription by recruiting the Sin3 repressor complex that contains

102260 64704860  
09840743 042301

catalytically active histone deacetylase (Jones *et al. Nature Genetics* 19(2):187-191 (1998)). Histone H3 and H4 deacetylation contributes to the formation of transcriptionally inactive chromatin. Thus, DMT can be used for the purpose of modulating the activity of target genes through chromatin architecture in animal cells as well as plant cells. For example, in some embodiments, DMT is used to catalytically remove 5-MeC from target gene DNA in several ways: e.g., (1) by fusing DMT to a sequence specific DNA binding protein, or (2) by fusing DMT to a subunit of the target repressor complex such as MeCP2 or Sin3. When combined with cell, tissue, or developmentally specific promoters DMT can be used to modulate specific sets of target genes.

In addition, reactive oxygen species, partially reduced species that are produced as intermediates of aerobic respiration, are powerful oxidizing agents that escape the mitochondria and attach to cellular components. Ionizing radiation and other agents that generate free radicals also produce reactive oxygen species that can attack the genome and cause lesions that are thought to have a key role in causing cancer and ageing. For example, 7,8-dihydro-8-oxoguanine (oxoG) is a very deleterious adduct generated by oxidation of the guanine base in DNA. The oxoG protein can pair with either cytosine or adenine during DNA replication. Thus, oxoG residues in DNA give rise to G/C to T/A transversion mutations. These transversions are common somatic mutations found in human cancers. HhH-GPD enzymes, such as those described herein, represent a defense against oxoG by catalysing the expulsion of the oxoG. Thus, in some embodiments, enhanced DMT activity is a method to reduce the incidence of mutations in animal cells. Also, DMT can be used to catalytically remove oxoG from a target gene by fusing DMT to a sequence specific DNA binding protein. When combined with a cell, tissue, or developmentally specific promoters DMT can be used to modulate repair of target genes.

As described above, the polypeptides of the invention can be targeted to chromosomal regions of interest by linking the polypeptides of the invention, including fragments with demethylase activity, to a DNA-binding domain that binds a target sequence. For example, it is known that an enzyme that methylates DNA (Dam methylase) can be targeted to specific sites in the genome (B.V. Steensel and S. Henikoff, *Nature Biotechnology* 18:424-428 (2000)). Specifically, the methylase was tethered to the DNA-binding domain of GAL4. When recombinant GAL4-methylase protein was expressed in transgenic *Drosophila*, targeted methylation occurred in a region of a few

09840743.042301

kilobases surrounding the GAL4 DNA binding sequence. In an analogous fashion, DMT, or a portion of DMT that has biological activity (e.g., a portion containing the HhH-GpD motif amino acids such as 1167 to 1368 of SEQ ID NO:2), can be tethered (e.g., as a translational fusion or chemically linked) to proteins that interact at specific sites in the genome. As a result, specific targeted regions of the genome are hypomethylated by DMT. As discussed above, typically hypomethylation promotes transcription of genes (S. E. Jacobsen, *Current Biology* 9, 617 (1999)). The invention provides compositions and methods for methylation of a desired area of the chromosome by targeting DMT to those regions. Thus, these embodiments provide additional ways to activate transcription of a desired gene in a targeted chromosomal region.

The following Examples are offered by way of illustration, not limitation.

#### EXAMPLE

##### **Example 1:**

This example shows the characterization of *dmt* mutant plants and the isolation of *DMT*.

Arabidopsis plants were transformed by infiltrating them with Agrobacterium containing the SKI15 T-DNA vector (generously provided by D. Weigel (Salk Institute, La Jolla, CA)). T1 seeds were harvested. The SKI15 vector has the bialaphos resistance (BAR) gene that allowed us to directly select transgenic plants in soil after spraying with the commercially available herbicide, Basta. Siliques from approximately 5,000 Basta resistant plants were opened, and those displaying approximately 50% seed abortion were identified.

Two lines, B13 and B33, were identified for further characterization. Genetic analysis of the mutants revealed that the *dmt* mutants were female sterile. Male fertility, however, depended on the genetic background of the mutant alleles. For instance, in the Columbia background, transmission of the *dmt* mutation is less than 50%. However, in the Landsberg *erecta* background, transmission through the male was almost normal.

Molecular analysis confirmed that the two mutations were allelic. For example, both the B13 and B33 alleles carry the SKI15 T-DNA within a DMT exon, confirming that disruption of the *DMT* gene resulted in the observed B13 and B33 phenotypes.



5'- and 3'- RACE were used to delineate the 5'- and 3'-ends of the cDNA, respectively. 5'-RACE was carried out using reagents and protocols provided by 5' RACE System for Rapid Amplification of cDNA Ends, Version 2.0, GIBCO BRL, LIFE TECHNOLOGIES, Grand Island, NY and Marathon cDNA Amplification Kit, Clontech, Palo Alto, CA. Final gene specific 5'-RACE primers were SKES-4 (GGGAACAAGTGACCATCTCC) and SKES3.5 (CGATGATACTGTCTCTTCGAGC). 3'-RACE was carried out using reagents and protocols provided by Marathon cDNA Amplification Kit, Clontech, Palo Alto. Final gene-specific 3' end was obtained from cDNA library screening.

The nucleotide sequence of the genomic copy of *DMT* was also determined (SEQ ID NO:1). The 5'-end of the *DMT* RNA is located at position 3,425 of SEQ ID NO:1. The position of the 3'-end of the *DMT* RNA is at position 12,504 of SEQ ID NO:1. The position of the ATG translation initiation codon is at position 4,903 of SEQ ID NO:1. The position of the TAA translation termination codon is at position 12,321 of SEQ ID NO:1.

A portion of the *DMT* polynucleotide sequence, including the first exon, is encompassed by the bacterial artificial chromosome (BAC) clone T9J15TRB. For example, sequences 3820-4299, 4319-4558, 4546-5025 and 9320-9777 of SEQ ID NO:1 were previously determined using the BAC clone as a template. Moreover, a separate independently sequenced region (Bork, C. *et al Gene* 28:147-153 (1998)) also overlaps the *DMT* sequence at positions 11,087 to 12,785 of SEQ ID NO:1.

The predicted *DMT* protein has 1,729 amino acids. This sequence was compared to known protein sequences using BLAST and revealed homology to several Endonuclease III proteins. The highest homology was to the Endonuclease III protein from *Deinococcus radiodurans*, Genbank Accession No. AE002073 (*see, e.g., White, O. et al. Science* 286:1571-1577 (1999)). Other *DMT* motifs include two consecutive nuclear localization signals at positions 43-60 and 61-78 and a leucine zipper at positions 1330-1351.

## **Example 2:**

This example provides further evidence that mutant phenotypes are caused by loss-of-function mutations.

A new allele, *dmt-3*, was obtained. The *dmt-3* allele was caused by insertion of the simple pD991 T-DNA vector (M. R. Sussman, *et al., Plant Physiol.*

124:1465 (2000)) into the 2nd exon of the *DMT* gene. In contrast, the previous two alleles, *dmt-1* and *dmt-2*, were caused by insertion of the activation T-DNA vector, SKI015 vector. The mutant phenotypes generated by all three *dmt* alleles are the same. Because pD991 does not have activation sequences, it suggests that all three mutant alleles are loss-of-function alleles. Consistent with this conclusion, seed abortion can be rescued with a transgene with 3,373 base pairs of 5'-*DMT* flanking sequences plus 1,478 base pairs of 5'-UTR ligated to a cDNA encoding the full-length DMT polypeptide (i.e., *DMTp::DMT*). Thus, *dmt/DMT* heterozygous plants that are hemizygous for the *DMTp::DMT* transgene displayed 25% seed abortion. Control *dmt/DMT* plants displayed 50% seed abortion.

### **Example 3:**

This example shows that *DMT* is necessary and sufficient for *MEA* gene expression.

As discussed above, when fertilization of *dmt/dmt* homozygous mutant flowers was prevented, fertilization-independent endosperm development was observed. This is very similar to when fertilization of mutant *mea* flowers is prevented. Thus, before fertilization, both *DMT* and *MEA*, a polycomb protein (T. Kiyosue *et al.*, *Proc. Natl. Acad. Sci. USA* 96:4186 (1999)), prevent the central cell of the female gametophyte from forming an endosperm. This is consistent with *DMT* being a positive regulator of MEDEA (*MEA*).

As further evidence of this relationship, *MEA* RNA accumulates in immature floral (IF) buds and open flowers (OF). However, in *dmt/dmt* mutant plants there was no detectable *MEA* RNA. Thus, *DMT* is necessary for *MEA* gene expression.

In addition, we have generated plants with a transgene, *CaMV::DMT*, designed to overexpress *DMT*. The full-length *DMT* cDNA was ligated to the constitutive cauliflower mosaic virus promoter, *CaMV* (S. G. Rogers, H. J. Klee, R. B. Horsch, R. T. Fraley, *Meth Enzymol* 153:253 (1987)). In control wild type plants, the *DMT* and *MEA* genes were not significantly expressed in the leaf. However, in *35S::DMT* plants, both *DMT* and *MEA* RNA level increased significantly. This shows that *DMT* is sufficient to induce *MEA* gene expression in the leaf.

#### Example 4:

This example shows that DMT is a member of the HhH-GPD superfamily of DNA repair enzymes.

A BLAST search, followed by a conserved domain search, revealed that DMT is highly related to the HhH-GPD superfamily of base excision DNA repair proteins (i.e., score of 70.1, E-value of  $8e^{-13}$ ). This family contains a diverse range of structurally related DNA repair proteins. The superfamily is called the HhH-GPD family after its hallmark helix-hairpin-helix and Gly/Pro rich loop followed by a conserved aspartate (S. D. Bruner, *et al.*, *Nature* 403:859 (2000)). This includes endonuclease III (EC:4.9.99.18), 8-oxoguanine DNA glycosylases (i.e., yeast OGG1), the thymine DNA glycosylase of methyl-CPG binding protein MBD4 (B. Hendrich, *et al.*, *Nature* 401:301 (1999)), and DNA-3-methyladenine glycosylase II (EC:3.2.2.21). The predicted amino acid sequence of DMT contains many of the conserved amino acids of this superfamily.

The hallmark of the superfamily of base-excision DNA repair proteins is a helix-hairpin-helix structural element followed by a Gly/Pro-rich loop and a conserved aspartic acid (i.e., HhH-GPD motif). The DMT polypeptide is 1,729 amino acids in length. Amino acids 1,271 to 1,304 correspond to the conserved HhH-GPD motif. The DMT sequence is DKAKDYLLSIRGLGLKSVECVRLTLHNLAFPVD. The catalytic lysine (K1286) and aspartic acid (D1304) residues are conserved in the HhH-GPD motif of DMT. Secondary structure prediction (Jpred program) indicates that DMT has two alpha-helices (amino acids 1,271 - 1,279 and 1,286 to 1,295) that correspond to the conserved alphaK and alphaL helices in the HhH-GPD motif of the crystallized hOGG1 DNA repair protein (Bruner *et al* *Nature* 403:859-866 (2000)).

The Arabidopsis *DMT* coding sequences were also used to identify homologous sequences in both public and proprietary databases using both the BLAST and PSI-BLAST computer algorithms. This analysis revealed amino acid sequences from several plant species, including wheat, maize, rice, soybean and Arabidopsis (SEQ ID NOs: 7-29). Based on these sequences, the following consensus sequences for DMT were determined:

#### DMT Domain A

KV<1>(I, l)D(D, p) (E, v) T<3>W<1>(L, v) L(M, l) (E, d) <0-  
2>D(K, e) <1>(K, t) <1>(K, a) (W, k) (W, l) <1>(E, k) ER<2>F<1>(G, t) R<1>(D, n) (S, l) FI (A, n) RM(  
H, r) <1>(V, l) QG(D, n) R<1>F<1>(P, q) WKGSVVDSV(I, v) GVPLTQN(V, t) D(H, y) (L, s) SS(S, n) A(F,  
Y) M<1>(L, v) A(A, s) <1>FP

## DMT Domain B

W(D,n)<1>(L,f)R<5>E<3>

6>D(S,t)<1>(D,n){Y,w}<3>R<10>I<2>RG(M,q){N,f}<2>L(A,s)<1>RI<2>  
12>FL<3>V<2>(H,n)G<1>IDLEWLR<2>(P,d){P,s}{D,h}<1>(A,v)K<1>(Y,f)LL(S,e){I,f}<1>G(L,i)GLKS(V,a)ECVRL<1>L(H,k)<2>AFPVDTNVGR(I,A,c)VR(M,l)G(W,l)VPL(Q,e)PLP<2>(L,v)Q(L,m)H(L,q)L(E,f)<1>YP<1>(L,m){E,d}{S,n){I,v}QK(F,y)LWPRCLK(L,p)Q<1>TYELHY(Q,h){L,m)ITFGK<0>2>FCTK<2>PNCNACPM(R,k)<0>2>EC(R,k){H,y}{F,y){A,s)SA<1>(A,v)<0>10>S(A,s){R,k)<1>(A,l)L(P,e)<1>(P,t)

## DMT Domain C.

P(I,l){I,v)E(E,f)P<1>(S,t)P<2>5>E<0>15>(D,a)IE(D,e)<4>

23>(I,v)P<1>I<1>(L,f){N,d}<8>17>(S,a)<1>(A,d)LV<8>(I,l)P<2>  
5>(K,r){L,m)K<4>LRTEH<1>V(Y,f){E,v)LPD<1>H<1>(L,i)L(E,k)<1>(D,e)D(P,i)<2>YLL(A,s){IW(T,q)P(G,d){E,g)<6>8>(P,s)<3>C<6>  
10>(M,l)C<4>C<2>C<3>(R,k)E<5>(V,f)RGT(L,i)L<0>  
22>(L,v)FADH<1>(S,t){S,r)<2>PI<3>(R,t)<3>(W,k)<1>L<1>(R,k)R<4>G(T,s){S,t)<2>(S,t){I(F,c){R,k){G,l)L<1>(T,v)<2>I<2>(C,n)F(W,q)<1>G(F,y){V,l)C(V,l)R<1>F(E,d)<3>(R,g)<1>P(R,k)<1>L<2>(R,h)LH<2>(A,v)SK

The first consensus sequence listed above corresponds to amino acid positions 586 through 937 of SEQ ID NO:2. The second consensus sequence listed above corresponds to amino acid positions 1117 through 1722 of SEQ ID NO:2. The consensus sequence provides amino acid sequences by position using single letter amino acid abbreviations. Numbers in carrots ("<" or ">") refer to amino acid positions where there is no consensus and which therefore, can be any amino acid. Amino acid abbreviations in parentheses indicate alternative amino acids at the same position. Capitalized letters refer to predominant consensus amino acids and lower case letters refer to amino acids that are commonly found in DMT sequences, but are not predominant.

### Example 5:

This example demonstrates the relationship between DNA repair and demethylation.

For many years, attention was focused on the ability of DNA glycosylases to repair DNA. For example, glycosylases are involved in the repair of G/T mismatched bases by depurinating the thymidine base moiety. Recently it was shown that avian (B.

09840745-042301

Zhu *et al.*, *Proc. Natl. Acad. Sci. USA* 97:5135 (2000)) and mammalian (B. Zhu *et al.*, *Nucl. Acid Res.* 28:4157 (2000)). G/T mismatch DNA glycosylases also have 5-methylcytosine-DNA glycosylase activity. That is, these enzymes are demethylases that remove 5-methylcytosine that is later replaced by cytosine. Without intending to limit the scope of the invention, it is believed that as a member of this superfamily, *DMT* is a demethylase (i.e., 5-methylcytosine glycosylase).

The methylation (i.e., amount of 5-methylcytosine) state of a gene can have a profound effect on its expression. In general, hypomethylation is associated with elevated gene expression, whereas hypermethylation is associated with decreased gene expression (S. E. Jacobsen, *Current Biology* 9:617 (1999)). Thus, it is likely that *DMT* activates *MEA* gene expression by reducing its level of methylation.

Mutations in the *DDM1* gene in Arabidopsis reduce by 70% the overall genome cytosine methylation (E. J. Finnegan, *et al.*, *Proc. Natl. Acad. Sci. USA* 93:8449 (1996); M. J. Ronemus, *et al.*, *Science* 273:654 (1996)). Such plants develop a number of phenotypic abnormalities including floral phenotypes (T. Kakutani, *et al.*, *Proc. Natl. Acad. Sci. USA* 93:12406 (1996)). Similarly, phenotypic abnormalities have been observed developing in *dmt/dmt* homozygous plants that affect petal number, floral organ fusion, and floral organ identity. Moreover, independent *CaMV::DMT* transgenic lines that overexpress *DMT* frequently are late-flowering. This is particularly interesting because late flowering of *ddm1* plants was shown to be due to hypomethylation of the *FWA* gene (W. J. J. Soppe *et al.*, *Mol Cell* 6:791 (2000)). Thus, without intending to limit the scope of the invention, it is believed that both *ddm1* loss-of-function mutations and overexpression of *DMT* (i.e., *CaMV::DMT*) may result in genome hypomethylation.

#### **Example 6:**

This example demonstrates targeting gene expression to the female gametophyte using a *DMT* promoter sequence.

*DMT* RNA accumulates in many plant organs such as immature flowers, mature flowers, open flowers, stems and to a lesser extent, leaves. To understand the spatial and temporal regulation of *DMT* RNA accumulation, the expression of the *DMT* promoter fused to reporter genes was analyzed. We fused 2,282 base pairs of 5'-*DMT* sequences, the full-length 5'-UTR (1,478 base pairs), 444 base pairs of *DMT* coding sequences that contain a nuclear localization signal to two reporter genes, the green

fluorescent protein (*GFP*; (Y. Niwa, *et al.*, *Plant J.* 18:455 (1999))) and  $\beta$ -glucuronidase (*GUS*; (R. A. Jefferson, T. A. Kavanagh, M. V. Bevan, *EMBO J* 6:3901 (1987))).

Reporter gene expression was observed in the developing female gametophyte, in the polar nuclei before they fuse, in the egg and synergids, and in the central cell. Expression  
5 was not detected after fertilization. Thus, this promoter is useful for targeting gene expression to the female gametophyte.

It is understood that the examples and embodiments described herein are for illustrative purposes only and that various modifications or changes in light thereof  
10 will be suggested to persons skilled in the art and are to be included within the spirit and purview of this application and scope of the appended claims. All publications, patents, and patent applications cited herein are hereby incorporated by reference in their entirety for all purposes.



CTTACTCAACAATTAGCCCAAAAAACATCCATAATGCATTTAACTAGGAATTTTAACAAAC  
 TCAAAATAGGTTGGTAGT  
 TAAAAAAAACAAATAGTAGATGTACATACGTACCTTTAAAAATATATACTCATATCGAAAGT  
 TTTAAATTTTGGCGAAAT  
 5 TAAATACATTATCTCATCAATTAATAATGCATAATTCTGTAATATCTATCTTT  
 AATTCCATATAGAAC  
 CAAAAAAAATAACATATCAAAATAGTTTAACTTAACAAAAACGTTAGGGAAAAGTTGACCT  
 AACTAGCTTGATTGACG  
 TTGAACCTGTCAATGCGAAAGCGATATTTCCAATATATACTACATGTAGTATTATTATATGGA  
 10 AGTTTCTAAAAAGGTG  
 TTGAGTGGATTGTTACTTGTGGAGGATGCTATTTTTCTCTCTGCCATAATTTTTACGAGTA  
 TGGGATAACTACATA  
 CTCATGATTATGAAACGCTCACTTTATTTGAAAAACCTCCTAATACACCAAAATATGCTACTAGA  
 TTCCAAAACGTAGACC  
 15 AATTGTATCTAATCTCAAAATCTCAATCAAAGTATTAATTTACCGATGGTAAGAAAAGTTAACC  
 GATATAATTATCAAAA  
 GAAAGAATAAGTCAACGAGATTCTAATCTCTTTATTTGGTATATGAACATTTGTACAAAAATC  
 TCAAAAGATATGTAAC  
 TGTTAAAAATATAAAATCACTGAGATTAATTCTTCAGACTCGTGTTAGCTATAAATGTCAAG  
 20 AGTTCTTCTGTTTCA  
 AGGAAAAACCTTAAAGATATGTATATTTCTGTAATTATGATGATATAATTTGCTATTCATTGT  
 CACAAACATTACTTTA  
 AAAAAATCGTATTTTACTTACTACAATGTTGACTAAGAACAAAAATACATTGATTATTGATATAT  
 CGTCAACTGAATTTTC  
 25 TTCCGAGGGATATAATTCTCAAACATAGCAAGAATCTCATAAATATGTTTCGTGACTACCTTTA  
 GACGAAATTTTTTTAA  
 GATTGTAACGTGACTTATGGTCTCTTGCTGTGGGGTCAATGCGAATAAATCTAAATGTATG  
 GGAGTCAAATAAAATAC  
 CAAGAAAAATAAGGAGCAGCACCCAATAAACTATATGGGACCAGAAATCCTTTTCATTGGTTT  
 30 AAAATAGGATTATCCCG  
 AAAGATGAAGGACTAAATTGAACTGATTGGGGGTAGGAAGAGATCCGTACAAATCATTAAAT  
 GGCTTCCACGCGGAAACT  
 TGTCGTTTATACAAATTCATTAACTTTCGGGTCGGGTTTATATTCCAAATGGGTCAAATAATAT  
 TAGTTTAATACACTAA  
 35 CGGAGTAATTAATTGGTGACTACAATTTATCAGTTTGGTGCAATTAGAAAACGAACATAGTCG  
 TAAAAACGAGTTCGGT  
 GTTATACCTTTATTTACGTTAAAAAATACGAGAATTTGTGTCAAATTTCAAATTAATTTTCAT  
 GAATATATGGAATAA  
 TTAGATACTCTAGCGAAAAATAGTATTATGAGCGTTTTACAAAAATACGATTTTAGCATTGAA  
 40 CTCCTTTATGTAATTC



GGTCAAAATGTTGGCATGAAGAAGCAAGTTTGCAACATTAAATTTTCATTTAAAAATCGTGTGA  
 CATACTTTAAAAATCTAA  
 ATATAGGAAGAAGACCAAAACATTAAATTTAGTAAGATTCTAATGAACATTATAAGGTATAA  
 CTTATAACCAACAAAAG  
 5 TTGGGTTTAGCGTGTGTTGCTTTATCTGAAAACCTGCAAACTAAACCATTTTAATAGGACTAATG  
 ACAATTAACAACAAA  
 TACACTTAAGCAACAACGTCCTCGTGAATATAATTTGGGCCTCAGGCCCATATTGCTAACGCC  
 AACTGATATTTCACITT  
 ATTCCTTCTTCATCTCACCACACTCTCTCTATCTCTATCTAACGGCATAGCTGACTCAGTG  
 10 TTCTCCGGCATTGAC  
 TCGCCTGAGAATCAGAAAGCTTAGATCGGTGAGCTTTTAGCTCCATTTTCTGTTTATTACATA  
 TTATTCTCTTTTCTC  
 TCTCTCCCTTTTATCTGGAAATTGTTCTGCTAAATTTCCAGCTGTTACATTTTCCGATCACG  
 AGAAGAATCACTGGG  
 15 TTTTATGTAAATCAATACATGTTCTGTTTCTGATCATAAATCTCAGCTATTAACACCTGATT  
 TTGATTCTGCGTAAT  
 AAAAACTCTGATTGCTTTTATCTTCACCTTTCCCATAAACATTGCTTACTTTATTGCTCTTC  
 TTTTACCGTTTCCAG  
 CTAATAAAATCTCTCGTATTCAATGTGTTCTCGTTTGTGTGATGAGAAAAATATCTGACAAAA  
 20 AATCAITTTATGCAIT  
 TTATGGTGACAGATTCTTAGTTAATGTCGCCTTCTCTAACCAAGTCAGATTAAAAAGGAGTGTTC  
 GTCCATGTTGCTTTGT  
 TTTGGTGTTTGGAGAGAGTTTTTCGGAGAGTTAGGTGAGTGTTATTTGGGGTGAGGTAGTGATA  
 AGGTTTGAAGGGGAGT  
 25 GATTTCATCAAGTGTGTTATGAATTCGAGGGCTGATCCGGGGGATAGATATTTTCAGATTCTTT  
 GGAGAATCAAACCTCAA  
 CAAGAGTTCATGGGTTCTTGATTCCATTACACCCAAAAAACCTAGATCAAGTCTGATGGTA  
 GATGAGAGAGTGATAAA  
 CCAGGATCTAAATGGGTTTCCAGGTGGTGAAATTTGTAGACAGGGGATTCTGCAACACTGGTGT  
 30 GGATCATAATGGGGTTT  
 TTGATCATGGTGCTCATCAGGGCGTTACCAACTTAAGTATGATGATCAATAGCTTAGCGGGAT  
 CACATGCACAAGCTTGG  
 AGTAATAGTGAGAGAGATCTTTTGGCAGGAGTGAGGTGACTTCTCCTTTAGCACCAGTTATC  
 AGAAACACCACCGTAA  
 35 TGTAGAGCCGGTCAATGGAAATTTTACTTCAGATGTGGGTATGGTAAATGGTCCTTTACCCA  
 GAGTGGCACTTCTCAAG  
 CTGGCTATAATGAGTTTGAATTGGATGACTTGTGAATCCTGATCAGATGCCCTTCTCCTTCAC  
 AAGCTTGCTGAGTGGT  
 GGGGATAGCTTATTCAGGTTCGTCATGTGAGTGATCAAATCTATTTTCAGTTTTTTTTTTTC  
 40 CTTTCTCCGTCTT

GCAGTACTTAGAGTAGAACATGAATTAGAATATCTTAAGAAAGTCATGGTTTTGAACAGATGG  
ACCTCCAGCGTGTAACA  
AGCCTCTTTACAATTTGAATTCACCAATTAGAAGAGAAGCAGTTGGGTCAGTCTGTGAAAGTT  
CGTTTCAATATGTACCG  
5 TCAAAGCCAGTCTGTTTCAGAACAGGTGAAAAGACTGGATTCTTGAACAGATAGTTACAAC  
ACTGGACATGAAATCCC  
AGAGCCGAAATCTGACAAAAGTATGCAGAGCATTATGGACTCGTCTGTGTTAATGCGCAGCGA  
AGCTACTGAACAAAATG  
ATGGCAGCAGACAAAGATGTTCTGGAGTTCGACCTTAACAAAACCTCTCAGCAGAAACCCCTCA  
10 AAAGGAAAAGGAAGTTC  
ATGCCAAAGGTGGTCTGTGGAAGGCAACCTAAAAGAAAGCCACGCAAACTGCAGAACTTCC  
CAAAGTGGTCTGTGGAAGG  
CAAACCTAAAAGGAAGCCACGCAAGCTGCAACTCAGGAAAAAGTGAAATCTAAAAGAAACCG  
GGAGTGCCAAAAAGAAA  
15 ATTTGAAAGAAATCAGCAACTAAAAAGCCAGCCAATGTTGGAGATATGAGCAACAAAAGCCCT  
GAAGTCACACTCAAAGT  
TGCAGAAAAGCTTTGAAATTTGACTTGGAGAATCCTGGAGATGCGAGGCAAGGTGACTCTGAG  
TCTGAAATTTGTCCAGAA  
CAGTAGTGGCGCAAACTCGTTTTCTGAGATCAGAGATGCCATTGGTGGAACTAATGGTAGTTT  
20 CCTGGATTCAAGTGTAC  
AAATAGACAAGACCAATGGATTGGGGGCTATGAACCAGCCACTTGAAGTGTCAATGGGAAAC  
CAGCCAGATAAACTATCT  
ACAGGAGCGAACTGGCCAGAGACCAACAACCTGATTTATTGACTAGAAACCAGCAATGCCA  
GTTCCCAAGTGGCAACCCA  
25 GAACACCCAGTTCCCAATGGAAAACCAACAAGCTTGGCTTCAGATGAAAAACCAACTTATTGG  
CTTTCCATTGTTGAAC  
AGCAACCTCGCATGACCATAAGAAACCAGCAGCCTTGCTTGGCCATGGGTAATCAACAACCTA  
TGTATCTGATAGGAACT  
CCACGGCTGCATTAGTAAAGTGGAAACCAGCAACTAGGAGGTCCCAAGGAAACAAGCGGCC  
30 TATATTTTGAATCACCA  
GACTTGTTTACCTGCTGGAAATCAGCTATATGGATCACCTACAGACATGCATCAACTTGTATG  
TCAACCGGAGGGCAAC  
AACATGGACTACTGATAAAAAACCAGCAACCTGGATCATTAAAGAGGCCAGCAGCCTTGC  
GTACCTTTGATTGACCAAG  
35 CAACCTGCAACTCCAAAAGGTTTTACTCACTTGAATCAGATGGTAGCTACCAGCATGTCATCG  
CCTGGGCTTCGACCTCA  
TTCTCAGTCACAAGTTCCTACAACATATCTACATGTGGAATCTGTTTCCAGGATTTGAATGGG  
ACTACAGGTACATGCC  
AGAGAAGCAGGGCTCTGCATACGATTCTTTACAGCAAGATATCCATCAAGGAAAATAAGTACA  
40 TACTTTCTCATGAGATA

TCCAATGGTAAATGGGTGCAAGAAAGCGTTACCTCAAAACTCTTCTCTGCCAACTCCAAATAATG  
GCTAAACTTGAGGAAGC  
CAGGGGCTCGAAGAGACAGTATCATCGTGCATGGGACAGACGGAAAGCATGATCTAAACT  
TAGCTCAACAGATTGCTC  
5 AATCAACAAGATGTGGAGAGACATAACAGCAGCAGGTGTGGAATATTTAGATGCTGCAAAAG  
AAAACGAAAAATCCAGAAA  
GTAGTCCAAGAAAAATTTGCATGGCATGCCACCTGAGGTTATAGAAATCGAGGATGATCCAACCT  
GATGGGGCAAGAAAAAGG  
TAAAAATACTGCCAGCATCAGTAAAGGTGCATCTAAAGGAAACTCGTCTCCAGTTAAAAAGAAC  
10 AGCAGAAAAAGGAGAAAT  
GTATTGTCCAAAAACGCCTGCAAAAAAGGGTCGAGCAGGTAGAAAAAAATCAGTACCTCCG  
CCTGCTCATGCCTCAGAG  
ATCCAGCTTTGGCAACCTACTCCTCCAAAGACACCTTTATCAAGAAGCAAGCCTAAAGGAAAA  
GGGAGAAAAGTCCATACA  
15 AGATTACAGGAAAGCAAGAGGTAACATAATGTATTCTACAATCTCTGTGATATAATTTTGAGAT  
TTTAGTAACGTGATGTGT  
CCAAACCAGCTCCTTATCACTGTGTGGTGCCTGTATAGGTCCATCAGGAGAACTTCTGTGTCAG  
GATTTCTATTGCGGAAA  
TAATTTACAGGATGCAAAATCTGTATCTAGGAGACAAAGAAAGAGAACAAGACAAAATGCA  
20 ATGGTCTTGTACAAAGGA  
GATGGTGCATTTGTTCCCTATGAGAGCAAGAAGCGAAAACCAAGACCCAAAGTTGACATTGA  
CGATGAACAACCTCGCAT  
ATGGAACCTTACTGATGGGGAAGGAGATGAAAAAGAGGGGATGAAGAGAAGGATAAAAAG  
AAAGAGAAGTGGTGGGAAG  
25 AAGAAAGAAGAGTCTTCGAGGAAGGGCTGATTCTTTCATCGCTCGCATGCACCTGGTACAAG  
GTGAAGATCCACTTCTC  
TTCTCAACTCCATTTTATTACACAAATTAGTAGAATACTCAAAAAATGATGTTTTGTTGCAA  
AATTTTAAAAATTCAT  
AGTTAACCATGTCAAAATAATTACATAATGCATCTTGTGAAGAACAGGTGTGCATTTATGGTG  
30 ACAGCTGAATGGTTTAT  
GTGCTATTATTTCTTTTACTGCTATAGATGACCAATTGAACTTAAACGTTTACAGGAGATAGA  
CGTTTTTCGCCATGGA  
AGGGATCGGTGGTTGATTCGGTCACTTGGAGTTTTCCTTACACAGAATGTCTCGGATCACCTTTC  
AAGTATATGAGTTGC  
35 CTTAATAAAATTGAGTTTCCAAAACATAGAAATTAACCCATGGTGGTTTACAATGCAGCTCTGC  
GTTTATGTCTCTAGCTG  
CTCGATTCCCTCCAAAATTAAGCAGCAGCCGAGAAGATGAAAGGAATGTTAGAAGCGTAGTT  
GTTGAAGATCCAGAAGGA  
TGCAATTCGAACTTAAATGAAATTCCTTCGTGGCAGGAAAAGGTTCAACATCCATCTGACATG  
40 GAAGTTTCTGGGGTTGA

TAGTGGATCAAAAGAGCAGCTAAGGGACTGTTCAAACTCTGGAATTGAAAGATTAAATTTCTT  
AGAGAAGAGTATTCAAA  
ATTTAGAAGAGGAAGTATTATCATCACAAAGATTCTTTTGATCCGGCGATATTTCAAGTCGTGTGG  
GAGAGTTGGATCTGT  
5 TCATGTTCCAAATCAGACGCAGAGTTTCCTACAACCAGGTGTGAAACAAAACTGTCAGTGGA  
ACATCACAAATCAGTGCA  
AACTGGGAGCCCAAACTTGTCTGATGAAATTTGTCTTCAAGGGAATGAGAGACCGCATCTATA  
TGAAGGATCTGGTGAT  
10 TTCAGAAACAAGAACTACAATGTCGCTCAGAAGAACTGATCTTGAAAAACAATGAAT  
TGGAAAGACTCTGTCTGT  
TTTGGTCAGCCAAGAAATGATACTAATTGGCAACAACCTCTCCAGCAGCTATGAGCAGTGT  
GCGACTCGACAGCCACA  
TGTACTAGACATAGAGGATTTTGGAAATGCAGGGTGAAGGCCCTTGGTTATTCTTGGATGTCCAT  
CTCACCAAGAGTTGACA  
15 GAGTAAAGAACAATAATGTACCACGCAGGTTTTTCAGACAAGGTGGAAGTGTTCGAAGAGAA  
TTCACAGGTCAGATCATA  
CCATCAACGCCTCATGAATTACAGGAATGGGATTGTCCGGTTCTCAAGCGCGTCCAAGAA  
CACCAGGACGATACCCA  
ACATAATCAACAAGATGAGATGAATAAAGCATCCCATTTACAAAAACATTTTGGATCTGCT  
20 CAACTCTCTGAAGAAT  
GCCTTACAAGACAGTCCAGTACCAACAGAATCATACGGATGGCTGTCTACCGAGAGATAGA  
ACTGCTGAAGACGTGGTT  
GATCCGCTCAGTAACAATTCAAGCTTACAGAACATAATTGGTCGAATCAAATTCAGCAATAAA  
GAGCAGACGGCAGTTGA  
25 ATACAAGGAGACAAATGCCACTATTTACGAGAGATGAAAGGGACGCTTGCTGATGGGAAAA  
AGCCTACAAGCCAGTGGG  
ATAGTCTCAGAAAAAGATGTGGAGGGGAATGAAGGGGAGACAGGAACGAAACAAAAACAATAT  
GGATTCCATAGACTATGAA  
GCAATAAGACGTGCTAGTATCAGCGAGATTTCTGAGGCTATCAAGGAAAGAGGGATGAATAA  
30 CATGTTGGCCGTACGAAT  
TAAGGTAATCTACTAATTTTCAGTTGAGACCCTCATCAAATCTGTCAGAAAGGCTTGAACATCA  
GTAAATTATGTAACCAT  
ATTTACAACATTGCAGGATTTCTAGAACGGATAGTTAAAGATCATGGTGGTATCGACCTTGA  
ATGGTTGAGAGAACTCT  
35 CTCCTGATAAAGCCAAGTGGGTAAATCACATTTTTAGTGACTGCAACACTAGCAGCATCGATT  
TACTCAACAATTACGTC  
AAACTGAGTATTACAAGTTGCTCATGAACATTTACAGGGACTATCTCTTGAGCATAAGAGG  
TCTGGGTTTGAAAAGTG  
TTGAATGCGTGCAGCTCTTAACACTCCA CAATCTTGCTTTCCCTGTGAGTCAGACTATTCCATT  
40 ATCTACTAAAACTTA

GAATAACTCCGGCTAACTAAGCTGGAACCTTGTATTGATGATATGAAGGTTGACACGAATGTTG  
GAAGGATAGCAGITAGG  
ATGGGATGGGTGCCTCTACAACCCCTACCTGAATCACTTCAGTTACACCTCCTGGAGCTGTAA  
GTTTCTTTTGTGTTGTC  
5 ATCTAAACAACGAAATTTTATGCAAGTCATAACCATGCTGTGTTTTCACAGATACCCAGTGTG  
CGAGTCCATCCAAAAA  
TTTCTTTGGCCAAGACTTTTGCAAACTCGATCAACGAACACTGTATGCTCATAAACTCTAACAAA  
TCATCTGTCTGAAAAA  
CCAATATTTCTTTGGTAGAATTCTATTGTCTATTACTACTAACAGCGAAAATTAATTAAACGT  
10 TCTTTTCTTACTCAG  
GTATGAATTACACTACCAACTGATTACGTTTGAAAGGTATTATTGCTCTAAGCTTTGAATTTA  
TCATATGGTAATTTCA  
AGCATTGTAGGCACCTGATCAATTATGTGTCTAAATCATGTGAATTCATGTCAGGTATTTTGCA  
CAAAGAGTAGACCAA  
15 TTGTAATGCATGTCCAATGAGAGGAGAGTGCAGACACTTTGCCAGTGCTTATGCTAGGTAAGC  
AAGCTTTCATGTACTTA  
TATGCAATAATTAAGATAAAATTTAGGATTATGGGTAAGTTACAAAAAATTAGGCTCAGTTT  
CATGGTAGCTAGCTGGA  
AATAGTATTACAAGAACAACATAAAGATCAAAGACAGAATCATGGATCCATATGCACTATCAT  
20 TTTAGCTCTTGTAAATCC  
ATACATGAACACTATATGCCAAAGTAGGGATTTCAAATATGAGATTTCGATGACTGATGCCATT  
GTAACAGTGCAAGACTT  
GCTTTACCGGCACCAAGGAGAGGAGCTTAACAAGTGCAACTATTCCGGTCCCTCCCGAGTCC  
TATCCTCCTGTAGCCAT  
25 CCCGATGATAGAACTACCTCTTCCGTTGGAGAAATCCCTAGCAAAGTGAGCACCATCGAATAG  
AGAAAAGTGTAACCAA  
TAATTGAAGAGCCGGCTCGCCCGGGCAAGAGTGCACTGAAATAACCGAGAGTGATATTGAA  
GATGCTTACTACAATGAG  
GACCCTGACGAGATCCCAACAATAAACTCAACATTGAACAGTTTGAATGACTCTACGGGAA  
30 CATATGGAAAAGAAACAT  
GGAGCTCCAAGAAGGTGACATGTCCAAGGCTTTGGTTGCTTTGCATCCAACAACACTTCTATT  
CCAAGTCCCAACTAA  
AGAACATTAGCCGTCTCAGGACAGAGCACCAGGTGAAGCTAATATCTCCTCTATATTTTATC  
TTCCATATAAAATTTG  
35 GGGAAAAAATCGCTCTCCATCTGGTTTTAGAACATGCGGGTCAGCCAGGGTTATGGCATTTTT  
ATATATTTACCGATCG  
GCCGAGCTGGCTCTGGTTGACTCGTATGCCACCCTGCATTGAACAAACAGTAGGAGACAA  
CAAGCAAAACGTTTTAA  
GATAAGGTCTATGGTAAATGACAAGGTAAGTATAAATGTGTCTCTATTGCAAGGTACGAG  
40 CTCCAGATTACATCG

TCTCCTTGATGGTGTAAGTCAATTTTAACTCTCTCTATACTCGAGTTGTTTCACTTGAGCAACA  
 CTGTTTAAAAGTCCT  
 CATTTGATAAAATAACAGATGGATAAAAGAGAACCAAGATGATCCAAGTCCTTATCTCTTAGCT  
 ATATGGACACCAGGTGA  
 5 GAATAAAACTGCAATGTTTCATTCATGTGTCTACAGTATCAAAGAAAGTACAGCTAGAGCTAA  
 AAAGCATTTGAAATAGA  
 GTCGGTTAAATATGAAAGTTTGAATCTGTAAATGAAAGCCGGAACGTAGCATTGGTGGATGTT  
 ATATGTAAATTAGTTTT  
 TGAGATTGGTCTAATGTAGTTGTTGACTGCCAGGTGAAACAGCGAATTCGGCACAACCGCCT  
 10 GAACAGAAGTGTGGAGG  
 GAAAGCGTCTGGCAAAATGTGCTTTGACGAGACTTGTCTGAGTGTAACAGTCTGAGGGAAGC  
 AAATCACAAGACAGTTC  
 GAGGAACCTCTCTGGTGAGATTATCTTGATCTTTTGTGTGTCTCATGAAAGGAGAAGTGAGA  
 ATACAAGTTTGCTAATA  
 15 TCATTTTTTCGTCATTCACAGATACCTTGTGCGACTGCCATGAGAGGAAGTTTTCCGCTCAACG  
 GGACATATTTCCAAGT  
 CAACGAGGTTAGATGAAATAAACTCAAACAGACAGACGAAACATTATTTCTGTTTAGTGTTG  
 GTCTTTATCCTCCTTG  
 CCATTTTTATCTTGCAGTTATTTGACAGACCAGAGTCCAGTCTCAAACCCATCGATGTTCTTA  
 20 GAGATTGGATATGGGA  
 TCTCCCAAGAGGACTGTTTACTCGGAACATCAGTAACATCAATATTCAGAGGTAAAAACAT  
 TCGTAATAGAGTTAGTT  
 AATCAAATGTCCAAAACAAGAAAGCTTCACCGTCCAATACACAAGAAAGCTTCACCTTCTC  
 TTTGCCAAAAAAGATCT  
 25 TAGAATGTTTGTGTAATTTGTGCAGGTCTTTCAACGGAGCAGATACAGTCTGCTTTTGAAAA  
 GGTAAACGTAACTTT  
 CGACCCAGAGAAATCCGAAAAATCTATTGCTTTGTTCTGATCAATACGTTAAACATATACACA  
 CACACTTTACACTTAGG  
 ACCAATACTGTTCTGATCTGTGATAGAACTGGTAAACATCTAACAATTATGATTGCAGGATT  
 30 CGTATGTGTCGTGGAT  
 TCGAACAGAAGACAAGAGCACCGCGTCCATTAATGGCAAGGTGTCATTTTCTGCGAGCAAAT  
 TGAAGAACAACAAAACC  
 TAAAGATGACTGGAAGAAAGCAAACGCATTGCTTCTGCTCTCTCTATTTAAAGCCAGGAA  
 AAGTCCCATTTAGACAT  
 35 AATAACAGGAATCCAATATAGGCTATTTTCTCTTCTTTCTTATTTCAATTCATAGAGCAGAAGCG  
 ACACAAAAAAGTTTTT  
 TGGGTATTTATTTTCTCTCTAACAATTTGTAGCGTTTGGGTCTTTTCTGGCTGTCACTAGC  
 GTGGCAATCCAATG  
 TCCGCGCACACTTAGGCGCATTGTCAATAAATCTCCGGCCACCGAGTGTTACGATCTTTTCC  
 40 AACGGCGGCTAATGCG

ATATTTCCGGTAACACATATTCCTTATTCTATGTTGGTTTTGTGTACGGCGTGGGCCTTACTAG  
ACAATGATCATCAATA  
AAACTAACACAAAGTTGAATGCTACAAAGTAGAAAGTGAAGAAAAATAATATAGACATTGC  
CGA

5

SEQ ID NO:2 DMT amino acid sequence

MQSIMDSSAVNATEATEQNDGSRQDVLFDLNLKTPQQKPSKRKRKFMPKVVEGKPKRKPRA  
ELPKVVVEGKPKRKP

10

KAAATQEKVSKETGSAKKKNLKESATKKPANVGDMNSKSPVTLKSCRKALNFDLENPGDARQG  
DSESEIVQNSSGANFS  
SEIRDAIGGTNGSFLDSVSQIDKTNGLGAMNQPLEVSMGNQPKLSTGAKLARDQQPDLLTRNQ  
CQFPVATQNTQFPME  
NQQAWLQMKNLQIGFPFGNQPRMTIRNQPPCLAMGNQQPMYLGITPRPALVSGNQQLGGPQGN  
KRPIFLNHQTCLPAGN

15

QLYGSPTDMHQLVMSTGGQGHLLIKNQPGSLIRGQQPCVPLIDQQPATPKGFTHLNMVATSM  
SSPGLRPHSQSQVPT  
TYLHVESVSRILNGTTGTCQRSRAPAYDSLQDDIHQGNKYLSHEISNGNGCKKALPQNSSLPTPIM  
AKLEEARGSKRQY

20

HRAMGQTEKHDLNLAQQIAQSQDVERHNSSTCEYDLAAKKTQKQVQENLHGMPEVIEIEDD  
PTDGARKGKNTASIS  
KGASKGNSSPVKKTAEEKCIVPKTPAKKGRAGRKSVPPAHASEIQLWQPTPPKTPLSRSKPKG  
KGRKSIQDSGKARG  
PSGELLQDSIAEIIHYRMQNL YLGDKEREQEQNAMVLYKGDGALVPYESKKRKRPRKVDIDDETR  
IWNLLMGKGDEKEG

25

DEEKDKKKKEKWWEEERRVFRGRADSFARMHLVQGDRRFSPWKGSVVDVSVIGVFLTQNVSDHLS  
SSAFMSLAARFPPKLS  
SSREDERNVRSVVVEDPEGCILNLNEIPSWQEKVQHPSDMEVSGVDSGSKEQLRDCSNSGIERFNL  
EKSIQNLEEEVLS  
SQDSFDPAIFQSCGRVSGSCSKSDAEFPPTTRCETKTIVSGTSQSVQTSQPNLSDEICLQGNRPHLYE

30

SGSDVQKQETTIN  
VAQKKPDLEKTMNWKDSVCFGQPRNDTNWQTPSSSYEQCATRQPHVLIEDFGMQGEGLGYS  
WMSISPRVDRVKKNVNP  
RRFFRQGGSVPREFTGQIIPSTPHELPGMGLSGSSAVQEHQDDTQHNQDEMNAKSHLQKTFLLDL  
LNSSEELTRQSSST

35

KQNITDGCLPRDRTAEDVDPLSNSSSLQNILVESNSSNKEQTAVEYKETNATILREMKGTADGK  
KPTSQWDSLTKDVE  
GNEGRQERNKNNMDSIDYEAIRRASISEISEAIKERGMNNMLAVRIKDFLERIVKDHGGIDLEWLRE  
SPDPAKADYLLSI  
RGLGLKSVECVRLTLHLNLAFFPVDTNVGRIVRMGWVPLQPLPESLQLHLELYPVLESIQKFLWP

40

RLCKLDQRTLYELH

YQLITFGKVFCTKSRPNCNACPMRGECRHFASAYASARLALPAPEERSLTSATIPVPPESEFPVVAIPM  
IELPLPLEKSLA  
SGAPSNRENCEPIIEEPASPGQECTEITESDIEDAYYNEDPDEIPTIKLNIEQFGMTLREHMERNMELQ  
EGDMSKALVAL  
5 HPTTTSIPTPKLNISRLRTEHQVYELPDSHRLLDGMDKREPDDPSYLLAIWTPGETANSAQPPEQ  
KCGGKASGKMCDF  
ETCSECNSLREANSQTVRGTLIPCRRTAMRGSFPLNGTYFQVNFELADHESLKPIDVPRDWIWDLP  
RRTVYFGTSVTSI  
FRGLSTEQIQCFCWKGFVCVRGFEQKTRAPRPLMARLHFPASKLKNNKT

SEQ ID NO:3 DMT 5' flanking sequence

AAGCTTAAAGCTACCAACATCGAATTTAGTAAAAGACCCATGATTTGAAATTGGAATTGTCGG  
CAAAATCGAGAAGATAT  
AGAGCCGACACGGGAACAGTGAAAACCAAAAGCGCGTAAGAAAGTGAACAGTGCGGAGAAGG  
15 AAGAGAGAATCTTACCGAT  
CATTCGAGGGAAAAGATGGGAATCAGAGAAAAATCTGGAAGAAAAAGAAATTAAGAGAAAAG  
GAGAGAAAGAAAGTGAGGAG  
GAAGATGCAGTGAAGACTGCTATAGCCACATCCACATGGTGTGATGAGAGAGAGAGAGAGA  
GAGGTTAAAGCAGCAAAAT  
20 TGTGGAGAGATAAAGAGAGAGAGAGACTGAGCGAGTCAAGTTCGTCGTCGTGTTAAAAAGAA  
AGAATCCTATATTTGCCT  
TTTTCTTTACTACTTTATTTTCAGACTATTTGCTTATTTTGCTCTCAAACCTTTTGTGATGTCACCT  
TTCGATCCTAAAGT  
GTTTGACAATTTACCTGCCTTTTCTCCAAGAAAAATCAGAACAGACCACAGCAAAATTTATGTA  
25 TTTTCTATTAAGAAAAAG  
AAAGAAAAGAAATTCATATTACTTATAGAAATTAAGCTAAGCAGTTGAAAACGTGAAAGCAGA  
ATTTCTAAAAAAATAGT  
AAACTGCTACAACTTATTTATGTGTATATAACATATCTATAAAGAACTCAAAATATATGATA  
AATCATTTTAACAAAAT  
30 TTCTATGAAATTATAATAAAAAAGTCACCTTTGACACTTAAAGGTTGACAATAACCGTCTCT  
CCAAAAAAATCAAA  
ACATTTATAATTTCTAAAACATGGTGTAATTTTGCTGAAATCAAAAAGAAAAGAAGGATTTC  
TATATCATAAGTTTCAT  
TATTGTATCAAACTTTCAAAATTTTCATGTAATTTGAAAGGAAAAAAATTAAGATATAATGTTGTT  
35 TTTGTTCTTATGTGA  
CATTTTCATGGAATATATATTCATAACAAAAATGTATTTTAATATGATGAGAGATTACCATCC  
AAAAGGTCGAACATTAT  
ATAAAACAAGTTAATACTAAACAATACATGTGATCACAATCAATGACAGTTTTGATCTTAA  
ATAGAAATGATTGAGCA



AACCTCAAAAATGCTCTTCTTAGGATCACAAAATCTTTCCTTAGCTTATTTAAAGCCGGGAGTTG  
AACTCTCTCTCCCTTG  
TAGACTTTTGTGTTTCAAATCTTTTTCTTTCAAAAAATCAATAATTAGTTAATTGGGCATAATATT  
TGGTTTTAATTAAGT  
CCATAGATTTTTTAGGACCATCTCTAATCACGACAAATATCTCAAATTGTAAACACATTTAAAC  
TTAAAAAGTATTGCATT  
CACAATCCTTAAAAATGAAAGTTAT  
ATAGAAACGATAA CTC  
CTTACTCAACAAATTAGCCCAAAAAAACATCCATAATGCATTTAACTAGGAATTTTAAACAAAC  
TCAAATAGGTTGGTAGT  
TAAAAAAAACAAATAGTAGTAGTACATACGTACCTTTAAAAATATATACTCATATCGAAAGT  
TTTAAATTTTGCGAAAT  
TAAATACATTTATCTATCAATTTAAATACATTTAATAATGCATAAATTCTGTAATATCTATCTTT  
AATTTCCATATAGAAC  
CAAAACAAAAATAAACATATCAAATAGTTTAACTTAACAAAAACGTTAGGGAAAAAGTTGACCT  
AACTAGCTTGATTGACG  
TTGAACCTTGTCATGTGCGAAAGCGATATTTCCAATATATACTACATGTAGTATTATTATATGGA  
AGTTTCTAAAAAGGTG  
TTGAGTGGATTGTTACTTGTGTGGAGGATGCTATTTTTTCCTTCTTGCCATAATAATTTACGAGTA  
TGGGATAACTACATA  
CTCATGATTATGAAACGCTCACTTTATTTGAAAAACCTCCTAATACACCAAATATGTCACTAGA  
TTCCAAAAACGTAGACC  
AATTGTATCTAATCTCAAATTTCTCAATCAAAGTATTAATTTACCGATGGTAAGAAAAGTTAAC  
GATATAATTATCAAAA  
GAAAGAATAAGTCAACAGATTCTTAATCTCTTTATTTTGGTATATGAACATTGTGCAAAAAATC  
TCAAAAGATATGTAAC  
TGTTTAAATATAAATTCACTGAGATTAAATCTTCAGACTCGTGTTAGCTATAATAATGTCAAG  
AGTCTCTCTGTTTCA  
AGGAAAAAACCTTAAAGATATGTATATTTTCTGTAATTATGATGATATAAATTGCTATTCTTGT  
CACAAACATTACTTTA  
AAAAATCGTATTTTCAATTACTACAATGTTGACTAAGAACAAAAATACATTGATTATGTATATAT  
CGTCAACTGAATTTTC  
TTCCGAGGGATATAATTCTCAAACATAGCAAGAATCTCATAATAATGTTTCGTGACTACCTTTA  
GACGAAATTTTTTAA  
GATTGCTAACGTGACTTATGGTCTCTTGCTGTGGGGGTCAATGCGAAATAAATCTAAATGTATG  
GGAGTCAAATAAAATAC  
CAAGAAAAATAAAGGAGCAGCACCCAATAAATATATGGGACCAGAAATCCTTTCATTGGTTT  
AAAATAGGATTATCCCG  
AAAGATGAAGGACTAAATTGAACTGATTGGGGGTAGGAAGAGATCCGTGACAAATCATTAAT  
GGCTTCCACGCGGAAACT

TGTCGTTTATACAATTTCATTAACITTCGGGTCGGGTTTATATTCCAAATGGGTCAAATAATAT  
 TAGTTTAATACACTAA  
 CGGAGTAATTAATTGGTGACTACAATTTTATCAGTTTGGTGCAATTAGAAAACGAACATAGTCG  
 TAAAAACGAGTTCGGT  
 5 GITATACCTTTATTTACGTTAAAAAATA CGAGAATTTTGTGTCAAATTTCAAATTAATTTTCAT  
 GAATAATGGAAATTA  
 TTAGATACTCTAGCGAAAAATAGTGATTATGAGCGTTTTACAAAAATACGATTTTAGCATTGAA  
 CTTCCTTTATGTAATTC  
 GGTCAAATGTGGCATGAAGAAGCAAGTTTGCAACATTAAATTTTCATTTAAAAATCGTGTGA  
 10 CATACTTTAAAAATCTAA  
 ATATAGGAAGAAGACCAAAACATTAAATTTAGTAAGATTCTAATGAACATTTATAAGTTATAA  
 CTTATAACCAACAAAAAG  
 TTGGGTTTAGCGTTGTTGCTTTATCTGAAAACTTGCAAACTAAACCATTTTAATAGGACTAATG  
 ACAATTAACAACAAAA  
 15 TACACTTAAGCAACAACGTCTCGTGAATATAATTTGGGCCTCAGGCCCATATTGCTAACGCC  
 AACTGATATTTCACTTT  
 ATTCCTTCTCATCTCACCACTCTCTCTATCTCTATCTCTAACGGCATAGCTGACTCAGT

20 SEQ ID NO:4 DMT 3' flanking sequence  
 AGATGACTGGAAGAAAGCAAAACGCATTGCTTCTCTGCTCTCCTCTATTTAAAGCCAGGAAAAAG  
 TCCCATTTAGACATAAT  
 AACAGGAATCCAAATAGGCTATTTTCTCTTCTTCTTATTTTCATTCTAGAGCAGAAGCGACA  
 CAAAAAAGTTTTTTGG  
 25 GTTATTATTTTCTCTCTAACAAAAAATAAAAAAAAAAAAAAACTCGAG

SEQ ID NO:5 DMT cDNA sequence  
 GTTCTCCGCATTGACTCGCCTGAGAATCAGAAAGCTTAGATCGGTGAGCTTTTAGCTCCATT  
 TCTGTTATTTACATA  
 30 TTATTTCTTTTTTCTCTCTCCCTTTTTATCTGGAATTTGTTCTGCTAAATTTCCAGCTGTTA  
 CATTTTCCGATCAC  
 GAGAAGAATCACTGGGTTTTATGTTAATCAATACATGTTCCTGTTTTCTGATCATAAATCTCA  
 GCTATTAACACCTGAT  
 TTTGATTCTGCGTAATAAAAAACCTCTGATTGCTTTTATCTTCACITTTCCCATAAACATTGCTT  
 35 ACTTTATTCGCTCTT  
 CTTTACCGTTTCCAGCTAAAAAATCTTCGCTATTCAATGTGTTTCTCGTTTGTGTGATGAGAA  
 AAATATCTGACAAAA  
 AATCATTTATGCAATTTATGGTGCAGATTCTAGTTAATGTCGCCITCTCTAACCAAGTCAGA  
 TTA AAAAGGAGTGTTC

GTCCATGTTGCTTTGTTTGGTGTGGAGAGAGTTTTCGGAGAGTTAGGTGAGTGTTATTGG  
 GGTGAGGTAGTGATAA  
 GGTTTGAAGGGGAGTGATTCATCAAGTGTGTTAATAATCGAGGGCTGATCCGGGGGATAGA  
 TATTTTCGAGTTCCTTT  
 5 GGAGAATCAAACTCAACAAGAGTTCATGGGTCTTGGATTCCATTTACACCCAAAAACCTAG  
 ATCAAGTCTGATGGTAG  
 ATGAGAGAGTGATAAACACGAGTCTAAATGGGTTCCAGGTGGTGAATTTGTAGACAGGGGA  
 TTCTGCAACACTGGTGTG  
 GATCATAATGGGGTTTTTGATCATGGTGCTCATCAGGGCGTTACCAACTTAAGTATGATGATCA  
 10 ATAGCTTAGCGGGATC  
 ACATGCACAAGCTTGGAGTAATAGTGAGAGAGATCTTTTGGGCAGGAGTGAGGTGACTTCTCC  
 TTTAGCACCAAGTTATCA  
 GAAACACCACCGTAATGTAGAGCCGGTCAATGGAATTTTACTTCAGATGTGGGTATGGTAA  
 ATGGTCCCTTACCCAG  
 15 AGTGGCACTTCTCAAGCTGGCTATAATGAGTTTGAATTGGATGACTTGTGAATCCTGATCAGA  
 TGCCCTTCTCCTTAC  
 AAGCTTGCTGAGTGGTGGGGATAGCTTATTCAAGGTCGCTAATGTGAGTGATCAAATCTATTT  
 TCAGTTTTTTTTTTC  
 CCTTCTCCGTTCTTGCACTTAGAGTAGAACATGAATTAGAATATCTTAAGAAAGTCATG  
 20 GTTTTGAACAGATGGA  
 CCTCCAGCGTGTAAACAGCCTCTTACAATTGAATTACCAATTAGAAGAGAAGCAGTTGGG  
 TCAGTCTGTGAAAGTTC  
 GTTTCAATATGTACCGTCAACGCCAGTCTGTTGAGAACAGGTGAAAAGACTGGATTCCCTGA  
 ACAGATAGTTACAATA  
 25 CTGGACATGAAATCCCAGAGCCGAAATCTGACAAAAGTATGCAGAGCATTATGGACTCGTCTG  
 CTGTTAATGCGACGGAA  
 GCTACTGAACAAAATGATGGCAGCAGACAAGATGTTCTGGAGTTCGACCTTAACAAAACCTCT  
 CAGCAGAAACCTCCA  
 AAGGAAAAGGAAGTTATCGCCCAAGGTGGTCGTGGAAGGCAAACTAAAAGAAAGCCACGCA  
 30 AACCTGCAGAACTTCCA  
 AAGTGGTCGTGGAAGGCAAACTAAAAGGAAGCCACGCAAGACTGCAACTCAGGAAAAAGTG  
 AAATCTAAAGAAACCGG  
 AGTGCCAAAAAGAAAAATTGAAAGAATCAGCAACTAAAAGCCAGCCAATGTTGGAGATAT  
 GAGCAACAAAAGCCCTGA  
 35 AGTCACACTCAAAAGTTGCAGAAAAGCTTTGAATTTTGAAGTGGAGATCCTGGAGATGCGAG  
 GCAAGGTGACTCTGAGT  
 CTGAAATTGTCCAGAACAGTAGTGGCGCAAACTCGTTTCTGAGATCAGAGATGCCATTGGTG  
 GAACTAATGGTAGTTTC  
 CTGGATTCAAGTGCACAAATAGACAAGACCAATGGATTGGGGGCTATGAACACGCCACTTGAA  
 40 GTGCAATGGGAAACCA

GCCAGATAAACTATCTACAGGAGCGAAACTGGCCAGAGACCAACAACCTGATTTATTGACTAG  
AAACCAGCAATGCCAGT  
TCCAGTGGCAACCCAGAACCCAGTTCCCAATGGAACCAACAAGCTTGGCTTCAGATGA  
AAAACCAACTTATTGGC  
5 TTTCCATTGTGTAACACAGCAACCTCGCATGACCATAAGAAACCAGCAGCCTTGCTTGGCCATG  
GGTAATCAACAACCTAT  
GTATCTGATAGGAACTCCACGGCTGCATTAGTAAGTGGAAACCAGCAACTAGGAGGTCCCCA  
AGGAAACAAGCGGCCTA  
TATTTTGAATCACCAGACTGTTTACCTGTGGAAATCAGCTATATGGATCACCTACAGACAT  
10 GCATCAACTTGTATG  
TCAACCGGAGGGCAACAACATGGACTACTGATAAAAAACCAGCAACCTGGATCATTAAATAAG  
AGGCCAGCAGCCTTGCGT  
ACCTTTGATTGACCAGCAACCTGCAACTCCTAAAGGTTTTACTCACTTGAATCAGATGGTAGCT  
ACCAGCATGTCATCGC  
15 CTGGGCTTCGACCTCATCTCTCAGTCACAAGTTCCTACAACATATCTACATGTGGAATCTGTTC  
CAGGATTTTGAATGGG  
ACTACAGGTACATGCCAGAGAAGCAGGGCTCCTGCATACGATTCTTTACAGCAAGATATCCAT  
CAAGGAAATAAGTACAT  
ACTTTCTCATGAGATATCCAATGGTAATGGGTGCAAGAAAGCGTTACCTCAAACTCTTCTCTG  
20 CCAACTCCAATTATGG  
CTAAACTTGAGGAAGCCAGGGGCTCGAAGAGACAGTATCATCGTGAATGGGACAGACGGAA  
AAGCATGATCTAAACTTA  
GCTCAACAGATTGCTCAATCACAAGATGTGGAGAGACATAACAGCAGCACGTGTGTGGAATA  
TTTAGATGCTGCAAAGAA  
25 AACGAAAATCCAGAAAGTAGTCCAAGAAAATTTGCATGGCATGCCACCTGAGGTTATAGAAA  
TCGAGGATGATCCAACCTG  
ATGGGGCAAGAAAAGGTAAAAATACTGCCAGCATCAGTAAAGGTGCATCTAAAGGAAACTCG  
TCTCCAGTTAAAAAGACA  
GCAGAAAAGGAGAAATGCTATTGTCCCAAAAACGCCTGCAAAAAAGGGTCGAGCAGGTAGAAA  
30 AAAATCAGTACTCCGCG  
TGCTCATGCTCAGAGATCCAGCTTTGGCAACCTACTCTCCAAAGACACCTTTATCAAGAAG  
CAAGCCTAAAGGAAAAAG  
GGAGAAAGTCCATACAAGATTCAGGAAAAGCAAGAGGTCCATCAGGAGAAGTCTGTGTGACG  
GATTCTATTGCGGAAAAA  
35 ATTTACAGGATGCAAAATCTGTATCTAGGAGACAAAGAAAGAGAACAAAGAGCAAAATGCAAT  
GGTCTTGTAACAAGGAGA  
TGGTGCACTTGTTCCTATGAGAGCAAGAAGCGAAAAACCAAGACCCAAAGTTGACATTGACG  
ATGAAACAACCTCGCATAT  
GGAACCTACTGATGGGAAAAGGAGATGAAAAAGAAGGGGATGAAGAGAAGGATAAAAAGAA  
40 AGAGAAGTGGTGGGAAGAA

GAAAGAAGAGTCTTCCGAGGAAGGGCTGATTCCCTCATGCTCGCATGCACCTGGTACAAGGA  
 GATAGACGTTTTTCGCC  
 ATGGAAGGGATCGGTGGTTGATTTCGGTCATTGGAGTTTTCTTACACAGAATGTCTCGGATCA  
 CCTTCAAGCTCTGGCT  
 5 TCATGTCTCTAGCTGCTCGATTCCCTCCAAAATTAAGCAGCAGCCGAGAAGATGAAAGGAATG  
 TTAGAAGCGTAGTTGTT  
 GAAGATCCAGAAGGATGCATTCTGAACTTAAATGAAATTCCTTCGTGGCAGGAAAAGGTTCAA  
 CATCCATCTGCACATGGA  
 AGTTTCTGGGGTTGATAGTGGATCAAAAGAGCAGCTAAGGGACTGTTCAAACCTCTGGAATTGA  
 10 AAGATTTAATTTCTTAG  
 AGAAGAGTATTCAAAATTTAGAAGAGGAAGTATTATCATCACAAGATTCTTTGATCCGGCGA  
 TATTTCAATCGTGTGGG  
 AGAGTTGGATCCTGTTTCATGTTCCAAATCAGACGCAGAGTTTCTACAACCAAGGTGTGAAACA  
 AAAACTGTCAGTGGAACT  
 15 ATCACAATCAGTGCAAACTGGGAGCCCAAACTTGCTGATGAAATTTGCTTCAAGGGAATGA  
 GAGACCGCATCTATATG  
 AAGGATCTGGTGATGTTTCAGAAAACAAGAACTACAAATGTCGCTCAGAAGAACTGATCTTG  
 AAAAAACAATGAATTGG  
 AAAGACTCTGCTGTTTTGGTTCAGCCAAGAAATGATACTAATTGGCAAAACAACCTCTCCAGC  
 20 AGCTATGAGCAGTGTGC  
 GACTCGACAGCCCATGTACTAGACATAGAGGATTTTGGAAATGCAAGGTGAAGGCCTTGTTA  
 TTCTTGGATGTCCATCT  
 CACCAAGAGTTGACAGAGTAAAGAACAAAAATGTACCACGCAGGTTTTTCAGACAAGGTGGA  
 AGTGTTCGAAGAGAATC  
 25 ACAGGTCAGATCATACCATCAACGCCTCATGAATTACCAGGAATGGGATTGTCGGTTCTCA  
 AGCGCCGTCCAAGAACA  
 CCAGGACGATACCCAACATAATCAACAAGATGAGATGAATAAAGCATCCCATTTACAAAAAA  
 CATTTTGGATCTGTCTA  
 ACTCCTCTGAAGAATGCCTTACAAGACAGTCCAGTACCAACAGAACATCACGGATGGCTGTC  
 30 TACCGAGAGATAGAAT  
 GCTGAAGACGTGGTTGATCCGCTCAGTAACAATTCAAGCTTACAGAACATATTGGTGAATCA  
 AATTCCAGCAATAAAGA  
 GCAGACGGCAGTTGAATACAAGGAGACAAATGCCACTATTTTACGAGAGATGAAAGGGACGC  
 TTGCTGATGGGAAAAAGC  
 35 CTACAAGCCAGTGGGATAGTCTCAGAAAAAGATGTGGAGGGGAATGAAGGGAGACAGGAACG  
 AAACAAAAACAATATGGAT  
 TCCATAGACTATGAAGCAATAAGACGTGCTAGTATCAGCGAGATTCTGAGGCTATCAAGGAA  
 AGAGGGATGAATAACAT  
 GTTGGCCGTACGAATTAAGGATTTCTAGAACGGATAGTTAAAGATCATGGTGGTATCGACCT  
 40 TGAATGGTTGAGAGAAT

CTCCTCCTGATAAAGCCAAGGACTATCTCTTGAGCATAAGAGGTCTGGGTTGAAAAAGTGTIG  
AATGCGTGCGACTCTTA  
ACACTCCACAATCTTGCTTTCCCTGTTGACACGAATGTTGGAAGGATAGCAGTTAGGATGGGA  
TGGGTGCCTCTACAACC  
5 CCTACCTGAATCAGTTACACTCCTCGAGCTATACCCAGTGCTCGAGTCCATCCAAAAA  
TTTCTTTGGCCAAGAC  
TTTGCAAACTCGATCAACGAACACTGTATGAATTACACTACCAACTGATTACGTTTGAAAGG  
TATTTTGCACAAAGAGT  
AGACCAAAATTGTAATGCATGTCCAATGAGAGGAGAGTGCAGACACTTTGCCAGTGCTTATGCT  
10 AGTGCAAGACTTGCTTT  
ACCGGCACCAGAGGAGAGGAGCTTAACAAGTGCAACTATTCCGGTCCCTCCCGAGTCCTTTCC  
TCCTGTAGCCATCCGA  
TGATAGAACTACCTTCTCCGTTGGAGAAATCCCTAGCAAGTGGAGCACCATCGAATAGAGAAA  
ACTGTGAACCAATAATT  
15 GAAGAGCCGGCTCGCCCGGGCAAGAGTGCAGTGAATAACCGAGAGTGATATTGAAGATGC  
TTACTACAATGAGGACCC  
TGACGAGATCCCAACAATAAACTCAACATTGAACAGTTTGAATGACTCTACGGGAACACAT  
GGAAAGAAACATGGAGC  
TCCAAGAAGGTGACATGTCCAAGGCTTGGTTGCTTTGCATCCAACAACACTACTTCTATTCCAAC  
20 TCCCAAACTAAAGAAC  
ATTAGCCGTCTCAGGACAGAGCACCAAGTGTACGAGCTCCAGATTACATCGTCTCCTTGAT  
GGTATGGATAAAAGAGA  
ACCAGATGATCCAAGTCTTATCTCTTAGCTATATGGACACCAGGTGAAACAGCGAATTCGGC  
ACAACCGCTGAACAGA  
25 AGTGTGGAGGGAAAGCGCTGGCAAAATGTGCTTTGACGAGACTTGTCTGAGTGTAACAGTC  
TGAGGGAAGCAAACTCA  
CAGACAGTTCGAGGAACCTTCTGTATACCTTGTGGAGTCCCATGAGAGGAAGTTTCCGCTC  
AACGGGACATATTTCCA  
AGTCAACGAGTTATTTGCAGACCACGAGTCCAGTCTCAAACCCATCGATGTTCTTAGAGATTG  
30 GATATGGGATCTCCAA  
GAAGGACTGTTTACTTCGGAACATCAGTAACATCAATATTAGAGGTCTTTCAACGGAGCAGA  
TACAGTTCTGCTTTTGG  
AAAGGATTCTGATGTGTCCGTGGATTGCAACAGAAAGACAAGAGCACCGCGTCCATTAATGGCA  
AGGTGTGCAATTTCTCTG  
35 GAGCAAAATTGAAGAACAACAAAACCTAAAGATGACTGGAAGAAAGCAAAACGATTGCTTCTC  
TGCTCTCCTCTATTAA  
GCCAGGAAAAGTCCCAATTTAGACATAATAACAGGAATCCAAATAGGCTATTTTCTCTTCTTTC  
TTATTTTCATTCATAGA  
GCAGAAGCGACACAAAAAAGTTTTTGGGTATTATTTTCTCTCTAACAAAAA  
40 AAAACTCGAG

SEQ ID NO:6 5' untranslated region of DMT

GTCTCCGGCATTGACTCGCCTGAGAATCAGAAAGCTTAGATCGGTGAGCTTTTAGCTCCATT  
TCTGTTATTACATA  
5 TTATTTCTTTTTTCTCTCTCCCTTTTTATCTGGAATTTGTTCTGCTAAATTTTCAGCTGT  
CAATTTCCGATCAC  
GAGAAGAACTCACTGGGTTTTATGTAAATCAATACATGTTCTGTTTCTGATCATAAATCTCA  
GCTATTAAACACCTGAT  
TTTGATTCTGCGTAATAAAAAACCTCTGATTGCTTTTATCTTCACITTTCCCATAAACAITGCTT  
10 ACTTTATTCGCTCTT  
CTTTTACCGTTTCCAGCTAAAAAATCTTCGCTATTCAATGTGTTTCTCGTTTGTGATGAGAA  
AAATATCTGACAAA  
AATCATTTATTGCATTTTATGGTGCAGATCTTAGTTAATGTCGCCTTCTCTAACCAAGTCAGA  
TTAAAAAGGAGTGTTCT  
15 GTCCATGTTGCTTTGTTTTGGTGTGAGAGAGTTTTCGGAGAGTTAGGTGAGTGTTATTTGG  
GGTGAGGTAGTGATAA  
GGTTTGAAGGGGGAGTGATTATCAAGTGTTATGAATTCGAGGGCTGATCCGGGGGATAGA  
TATTTTCGAGTTCCTTT  
GGAGAATCAAACTCAACAAGAGTTTCATGGGTCTTGGATTCCATTTACACCAAAAAACCTAG  
20 ATCAAGTCTGATGGTAG  
ATGAGAGAGTGATAAAACAGGATCTAAATGGGTTTCCAGGTGGTGAATTTGTAGACAGGGGA  
TTCTGCAACACTGGTGTG  
GATCATAATGGGGTTTTTGATCATGGTGCTCATCAGGGCGTTACCAACTTAAGTATGATGATCA  
ATAGCTTAGCGGGATC  
25 ACATGCACAAGCTTGGAGTAATAGTGAGAGAGATCTTTTGGGCAGGAGTGAGGTGACTTCTCC  
TTTAGCACCAGTTATCA  
GAAACACCAACCGGTAATGTAGAGCCGGTCAATGGAAATTTTACTTCAGATGTGGGTATGGTAA  
ATGGTCTTTACCCAG  
AGTGGCACTTCTCAAGCTGGCTATAATGAGTTGAATGGATGACTTGTGAATCCTGATCAGA  
30 TGCCCTTCTCCTCAC  
AAGCTTGTGAGTGGTGGGGATAGCTTATTCAAGGTTCTGCAATGTGAGTGATCAAATCTATTT  
TCAGTTTTTTTTTTC  
CCTTCTTCGGTCTTGCGTACTTAGAGTAGAACATGAATTAGAATATCTTAAGAAAGTCATG  
GTTTTGAACAGATGGA  
35 CCTCCAGCGTGTAAACAAGCCTCTTACAATTTGAATTCACCAATTAGAAGAGAAGCAGTTGGG  
TCAGTCTGTGAAAGTTC  
GTTTCAATATGTACCGTCAACGCCAGTCTGTTCAGAACAGGTGAAAAGACTGGATTCTTGA  
ACAGATAGTTACAATA  
CTGGACATGAAATCCCAGAGCCGAAATCTGACAAAAAGT

(gi 7406444);

64441 tcactagatt ccaaaacgta gaccaattgt atctaattctc aaattctcaa tcaagtgatt  
64501 aatttacoga tggtaagaaa agttaacoga tataattatc aaaaagaaga ataatgcaac  
64561 agattcttaa tctctttatt ttggtatgat aacatttgta caaaaatctc aaaaagatatg  
64621 taactgttta aaatataaat tcactgagat taattcttca gactcgtgtt agctataata  
64681 atgccaagag tctctctgtt ttcaagggaaa aaccttaaa atagtatat ttctgttaat  
64741 tatgatgata taatttgcta ttcattgtca caaacattac tttaaaaaat cgtattttca  
64801 ttactacaat gttgactaag aacaaaaata cattgattat tgatatatcg tcaactgaat  
64861 ttctctcoga gggatataat tctcaaacat agcaagaatc tctaataat gttcgtgac  
64921 tacccttaga cgaatttttt ttaagattcg taactgtact tatggtctct tgcctgtggg  
64981 gtcaatgoga ataatctaa atgtatggga gtcaaataaa ataccaaga aataaaggga  
65041 gcagcaccca ataaactata tgggaccaga aatcctttca ttggtttaaa ataggattat  
65101 ccogaagaat gaaggactaa attgaaactg attgggggta ggaagagatc cgtcaaatc  
65161 attaatggct tccacgggga aacttgtcgt ttatacaatt tcaattaaact tcgggtcggg  
65221 ttatatattcc aatgggtgca aataatatta gtttaataca cttaaaggagt aattaaattgg  
65281 tgactacaat tttatcagtt tgggtgcaat agaaaacgaac atagtcgtaa aatacagatt  
65341 cgggtgttata cctttattta cgttaaaaaa atacgagaat ttgtgttcaa atttcaaat  
65401 aatttcatga atatatggaa atatttagat actctagcga aataatgtat tatgactgat  
65461 ttacaaaaat acgatttttag cattgaacct cctttatgta atcggtcaa atgttggcat  
65521 gaagaagcaa gtttgaaca ttaaatctca tttaaaaatc gtgttgacat actttaaaaat  
65581 ctaaatatag gaagaagacc aaaaacttaa atttagtaag attctaata cattaataat  
65641 gttataaact ataaccaaca aagttgggt ttacgctgtg tgccttatct gaaaacttgc  
65701 aaactaacc attttaatag gactaatgac aattaacac aaaaatacat tgaacaacaa  
65761 cgtcctcgtg aataataatt gggcctcagg cccatattcg taacgcacac tgatatttca  
65821 ctttatctct tcttcactct accacactct ctctctatct ctatctctaa cggcagatct  
65881 gactcagtgat tctccggcat tgactcgtcg gagaatcaga aagcttagat cgggtgagct  
65941 tagctcccat ttctctgtta ttactaat atttctcttt ttctctctct ccttttttat  
66001 ctgggaatttg ttctgtctaaa ttttccagct gttacatttt ccgatacaga gaagatcac  
66061 tgggtttttaa tgttaatacaa tacatgttcc tgttttttga tctaataatc cagctattaa  
66121 cacctgattt tgattctcgt taataaaaaa cctctgattg cttttatctt cactttcccc  
66181 ataaacattg ctactcttat tgcctctctc ttacccgttt ccagctaaaa aattctctgc  
66241 tattcaatgt gttctcgttt ttgttgatga gaaaaaatatc tgacaaaaaa tcaattattg  
66301 cattttatgg tgcagattct tagttaatgt gcctctctct aaccaagctca gattaaaaag  
66361 gagtgctcgt ccattgtgct ttgttttggg gtttgagag agttttcga gatttaggtg  
66421 agtggttattt ggggtgaggt agtgataagg ttgaaagggg gaggattca tcaagtgctg  
66481 tatgaattcg agggctgac cgggggatag atatttttga gttcctttgg agaataaac  
66541 tcaacaagag ttcatggggt ctgggatcc atttaccac aaaaaacctc gatcaagctg  
66601 gatggttagat gagagagtga taaaccagga tctaaatggg ttccagggtg tcaagtttgt  
66661 agacagggga ttctgcaaca ctggtgtgga tcaaatggg gtttttgatc atggtgctca  
66721 tcagggctgt accaacttaa gtatgatgat caatagctta gcgggatcac atgcaaacg  
66781 ttggagtaaat agtgagagag atcttttggg caggagtgag gtactcttc ctttagcacc  
66841 agttatcaga aacaccacog gtaatttaga gcgggtcact ggaattttta cttcagatg  
66901 gggatgggta aatggtctct tccaccagag tggcactctc caagctggct ataatgattg  
66961 tgaattggat gacttggtga atcctgatca gatgcccttc tctctcaaca gcttgctgag  
67021 tgggtggggt agcttattca aggttcgtca atgtgagtga tcaaatctat tttagtttt  
67081 ttcttttccc ttctctcgtt ccttgagta cttagagtag aacatgaatt agaatattct  
67141 aagaagtgca tgggtttgaa cagatggacc tccagctgtg aacaagcttc tttaaatct  
67201 gaattcacca attagaagag aagcagctgg gtcagctctg gaaagtcgt ttcaatatg  
67261 accgtcaacg cccagctctg tcgaacacag tgaaagagct ggaattctgt aacagatag  
67321 tacaactact ggacatgaaa tccagagacc gaaatctgac aaaaagtATGc agagcattat



50

64

5

10

15

20

25

30

35

40

45

50

70441 tcgtgtggga gagttggatc ctgttcatgt tccaaatcag acgcagagtt tcttacaacc  
70501 aggtgtgaaa caaaaactgt cagtggaaca tcacaatcag tgcaaatcgg gagcccaaac  
70561 ttctgtcatg aaatttgtct tcaagggaat gagagacgc atctatata aggatctggt  
70621 gatgttcaga aacaagaaac tacaatgtc gtcagaaga aacctgatct tgaaaaaaca  
70681 atgaattgga aagactctgt ctgttttggg cagccaagaa atgatactaa ttggcacaac  
70741 actccttcca gcagctatga gcaggtgtgc actcgacagc cacatgtact agacatagag  
70801 gatttttgaa tgcaagggtga aggccttggg tatctttgga tgtccatctc accaagagtt  
70861 gacagagtaa agaacaaaaa tgtaccacgc aggtttttca gacaagggtg aagtgttcca  
70921 agagaattca caggtcagat cataccatca acgcctcatg aattaccagg aatgggattg  
70981 tcgggttctc caagcgcctg ccaagaacac caggagcaga ccaacataaa tcaacaagat  
71041 gagatgaata aagcatccca tttaaaaaa acatttttgg atctgtctaa ctctctgaa  
71101 gaatgcctta caagacagtc cagtacccaa cagaacatca cggatggctg tctacogaga  
71161 gatagaactg ctgaagacgt ggttgatcgg ctacgtaaac attcaagctt acagaacata  
71221 ttggtcgaat caaatccag caataaagag cagacggcag ttgaatacaa ggagacaaat  
71281 gccactattt tacgagagat gaaggggagc ctgtctgatg ggaaaaagcc tacaacccag  
71341 tgggatagtc tcagaaaaa gttggggggg aatgaaggga gacagggaac aaacaaaaac  
71401 aatatggatt ccatagacta tgaagcaata agacgtgcta gtatcagcga gattttctgag  
71461 gctatcaagg aagaggggat gaataacatg ttggcctgac gaattaaggt aaactctacta  
71521 atttcagctg agacccctcat caaatctgtc agaaggcttg aacatcagta aattatgtaa  
71581 ccatatttac aacattgcag gatttctcat aacggatagt taaagatcat ggtggatcga  
71641 accctgaatg gttgagagaa ttctctcctg ataaagccaa gtgggttaaa cacattttta  
71701 gtgactgcga cactagcacg atcgatttac tcaacaatta cgtcaaacgt agtattaaac  
71761 agttgctcat gaacatttca cagggaactat ctcttgagca taagagggtct ggggttgaaa  
71821 agtgttgaaat gcgtgcgact cttaacactc cacaatcttg ctctccctgt gactcagact  
71881 attccattat ctactaaaaa cttagataaa ctccggctaa ctaagctgga actgtattg  
71941 atgatatgaa ggttgacacg aatgttgga ggatagcagt taggatggga ttgggtgctc  
72001 tacaacccct accgtaatca ctctcagttac accctctgga ctgtgaagt ttcttttgtt  
72061 tgtcatctaa acaacgaaat ttttatgcga gtcatcaacca tgtgtgttt tcacagatac  
72121 ccagtgctcg agtccatcca aaaaatttct tggccaagac ttgcaaaact cgaatacga  
72181 acactglatg ctcataaaat cttaacaaat atctgtctga aaaaaccaat ttcttttggt  
72241 agaattctat tgtcattact cattactaac agcgaatta attaacgttc ttctttctac  
72301 tcagggtatga attacactac caactgatta cgtttggaaa ggtattattg ctctaagctt  
72361 tgaatttata atatggtaat ttcaagcatt gtaggcacct gatcaattat gtgtctaaat  
72421 catgtgaatt catgtcaggt attttgaca aagagtagac caaattgtaa tgcattgtcca  
72481 atgaggagg agtgcgagaca ctttgcaggt gcttatgcta ggtgaagcag ctttatgtga  
72541 ctttatatga ataattaaag ataaaaatta ggattattgg taagtaacaa aaaattaggc  
72601 tcagtttcat ggtagctagc tggaaatagt attacaagaa caacataaag atcaacagca  
72661 gaactcatgga tccatattga ctatcatttt agctcttgta atccatcac gaacataata  
72721 tgccaaagta gggatttcaa atatgagatt cgtagactga tgcatttga acagtgcag  
72781 actgtgttta ccggaccacg aggagaggag cttaacaagt gcaactattc cggctccctc  
72841 cagtcctat cctcctgtag ccatcccgat cttagaacta cctctccgt tggagaactc  
72901 cctagcaagt ggagcaccat cgaatagaga aaactgtgaa ccaataattg aagagccggc  
72961 ctgcgccggg caagagtga ctgaataaac cggagatgat attgagagat ctactacaaa  
73021 tggagacct gacgagatcc caacaataaa actcaacatt gaacagcttg gaactgact  
73081 acgggaacac atggaaagaa acatggagct ccaagaaggt gacatgtcca aggctttgg  
73141 tgccttgcac ccaacaacta ctctctatcc aactcccaaa ctaagaacaa ttacgcgtct  
73201 caggacagag caccaagtgt aagcataat ctctctatct attttattc ccatataaat  
73261 tttggggaaa aaactcgtct ccatctgggt ttagaacatg cgggtcagcc agggttatgg  
73321 catttttata tatttccacc atcgcccgga gctggctctg gttgactcgt atggccacct  
73381 gcattgaaca aaccagtagg agacaagcaa gcaaaaagct ttaagataag gttcattgta  
73441 aaatgacaag gtcaactgata aatgtgtcgt ctattttgag gtacagagtc ccagattcac

5

10

15

20

25

73501 atcgtctcct tgcgtggtga agtcaatttt taactctctc tatactcgag ttgtttcaact  
 73561 tgagcaacac tgtttaaaag tctctatttg ataaaaaac agatggataa aagagaacaa  
 73621 gatgatccaa gtctttatct cttagctata tggacaccag gtgagaataa aactgcaatg  
 73681 ttctattcat gtgtctacag tatcaaaaga agtacagcta gagctaaaaa gcatttgaaa  
 73741 tagagtcggg taaatatgaa agtttgaac tgtaaatgaa agccggaaag tagcattggg  
 73801 ggatgttata tgtaaattag tttttgagat tgggtctaag tagttgtttg actgccaggt  
 73861 gaaacagcga attcgggcaca accgcctgaa cagaagtgtg gagggaaaag gtctggcaaa  
 73921 atgtgctttg acgagacttg ttctgagtgt aacagcttga ggggaagcaa ctccacagaca  
 73981 ggtcgaggaa ctcttctggg gagattatct tgatcttttg tgttgcctat gaaaaggaga  
 74041 agtgagaata caagtittgct aatatcattt ttctgtcatt cacagatacc ttgtcggact  
 74101 gccatgagag gaagttttcc gctcaacggg acatatctcc aagtcaacga ggttagatga  
 74161 aataaaaactc aaacagacag acgaaacatt atttctgttt agtgttggtt ctttatcttc  
 74221 cttgccattt ttatctttgc agttatttgc agaccaagag tccagtctca aacctatcga  
 74281 tgttctctaga gattggatat gggatctccc aagaaggact gtttacttcg gaacatcagt  
 74341 aacatcaata ttcagaggta aaaacattcg taatagagtt agttaatcaa atgtccaaaa  
 74401 cacaagaaga cttcacogtc caatacacaa gaaagcttca cttctcttt gccaaaaaag  
 74461 atcttagaat gttttgctga atttgtgcag gtctttcaac ggagcagata cagttctgct  
 74521 ttggaaaagg taaacgttaa ctttcgaccc agagaaatcc ggaaaatcta ttgctttggt  
 74581 ctgatcaata cgttaaacat atacacacac actttacact taggaccaat actgttctga  
 74641 tctgtgatag aaactggtaa acatctaaca attatgattg caggattcgt atgtgtccgt  
 74701 ggattcgaac agaagacaag agcacccggt ccattaatgg caaggttga tttctctcgc  
 74761 agcaaatga agaacaacaa aacctAAaga tgactggaag aaagcaaacg cattgtctct  
 74821 ctgctctcct ctattttaaag ccaggaaaag tcccatttag acataataac aggaatccaa  
 74881 ataggctatt ttctctttct ttcttatttc attcatagag cagaagcgac acaaaaaagt  
 74941 tttttgggtt atttatttct tctctaacia atttgtagcg ttttgggtct ttttctggct  
 75001 gtcactagcg tggcaaatcc aatgtctcgc cacacttagg cgcattgtca ataaaaatttc

SEQ ID NO:8

ARABIDOPSIS THALIANA DMT2

>DMT2 (1DMT2) ;

MEKQRREBESSFQQPPWIPQTPMKPFSPICPYTVEDQYHSSQLEERRFVGNKDMSGLDHLS  
5 FGDLLALANTASLIFSGQTPIPTRNTEVMQKGTTEEVESSLSSVNNVAEQILKTPEKPKRK  
KHRPKVRREAKPKREPAPRPRKSVVTDGQESKTPKRKYVRKKVEVSKDQDATPVESAA  
VETSTRPKRLCRRVLDFAEANGENQTNQDIREAGEMESALQEQQLDSGNQELKDCLLSAP  
STPKRKRSGQKRKGVPKKNKSNLEEVDISMAQAARRQGPTCCDMNLSGIQYDEQCDYQ  
KMHWLYSPNLQQGGMYDAICSKVFSGQQHNYVSFAFHATCYSSSTSQLSANRVLTVEERRE  
10 GIFQGRQSEELNVLSKIDTPIKKTTHGARFRNLSSMNKLVEVPEHLTSGYCSKPQQNN  
KILVDTRVTVSKKPKTKSEKSTQKQKNLLPNLCRFPPSFTGLSPDELWKRNSIETISEL  
LRLLDINREHSETALVPYTMNSQIVLFGGGAGAIVPVTVPKKPRPRPKVLDLDETRVWK  
LLENINSEGVDGSDQBQAKWEEERNVFRGRADSFIAARMHLVQGGRRFTPWKGSVVDVS  
VGVFLTQNVSDHLSSSAFMSLASQFPVPFVPSNFDAGTSSMPSIQITYLDSEETMSPPP  
15 DHNHSSVTLLKNTQPDEEKDYVPSNETSRSSSEIAISAHESVDKTTDSKEYVDSDRKGSSV  
EVDKTDCECRVLNLFPSSEDSALTCQHSVMVSDAPQNTERAGSSSEIDLEGEYRTSFMKLLQ  
GVQVSLEDSNQVSPNMSPGDCSSSEIKGFQSMKEPTKSSVDSSEPGCCSQQGDVLSQCQP  
TLKEKGKVLKKEKKAFFDWDCLRREAQARAGIREKTRSTMDTVDWKAIRAADVKEVAETI  
KSRGMNHKLAERIQYLTLMNKIMQGFLDRLVNDHGSIDLEWLRDVPPDKAKEYLLSFNGL  
20 GLKSVCEVRLLTLHLHAFVPDNTVGRIAVRLGWVPLQPLPESLQLHLEMPMLESIQKY  
LWPRCLKLDQKTLTYELHYQMITFGKVFTCKSKPNCNACPMKGECRHFAFAFARKFSNIHL  
FYSARLALPSTTEKGMGTPDKNPLPLHLPEPFQREQGSEVVQHSEPAKVTCCEPIIEEPA  
SPEPETAEVSADIIEEAFFEDPEEIIPTIRLNMDAFTSNLKKIMEHNKELQDGNMSSALVA  
LTAETASLPMPKLKNISQLRTEHRVYELPDEHPLLAQLEKREDDPCSYLLAIWTPGETA  
25 DSIQPSVSTCIFQANGMLCDEETCFSCNSIKETRQIVRGITILIPCRATAMRGSPFLNGTY  
FQVNEVFADHASSLNIPINVPRELIWELPRRTVYFGTSSVPTIFKGLSTEKIQACFWKGYVC  
VRGFDRKTRGPKPLIARLHFPASKLKGQANLA

SEQ ID NO:9

>DMT2 (1DMT2) novel 480 amino acid amino terminus;

MEKQRREBESSFQQPPWIPQTPMKPFSPICPYTVEDQYHSSQLEERRFVGNKDMSGLDHLS  
30 FGDLLALANTASLIFSGQTPIPTRNTEVMQKGTTEEVESSLSSVNNVAEQILKTPEKPKRK  
KHRPKVRREAKPKREPAPRPRKSVVTDGQESKTPKRKYVRKKVEVSKDQDATPVESAA  
VETSTRPKRLCRRVLDFAEANGENQTNQDIREAGEMESALQEQQLDSGNQELKDCLLSAP  
STPKRKRSGQKRKGVPKKNKSNLEEVDISMAQAARRQGPTCCDMNLSGIQYDEQCDYQ  
35

KMHWLYSPNLQQGGMRYDAICSVFSGQQHNVVSAPFATCYSSTSQLSANRVLTVEEERRE  
GIFQGRQSESLNVLSDKIDTPIKKKTTGHARFRLSSMNKLVEVPEHLTSGYCSKPPQQNN  
KILVDTRVTVSKKKPTKSEKSTQTKQKNLLPNLCRFPSPFTGLSPDELWKRNRNSETISEL

# 5 SEQ ID NO:10

>DMT2 (1DMT2) Nucleotide sequence from BAC F1011 (gi 6598632);

60001 tgcgtgagcc tgggtttctt catcggaact ggatctctgg atctatcaaa cgggtctacga  
60061 ggattctcca ttccaagaaa ctatacaata caagaggtag gcaaat aatg cccataatta  
60121 aacctaatcg gcaaaaaatcg attgcagtag caacaaatcc tegttagagg ggaattcaga  
60181 gcaattacaac aatcagtaac cctaagttac aatctaaaaa ttgagatgca taacgagatt  
60241 ctgcgaagaa gacggagaag atagaaggaa tgcttcgaat tcggcaaaaa tgcagagag  
60301 ttgggacaat ctccgatcaa ttgggtttgt gaattgggga ttttatggag acgagacaaa  
60361 aaaaagttag agatcggagc tgggttccaa aatatttagg cccatttaac gaccocattt  
60421 ccatgtataa taggcgccatc atctaataat tgacaaacat agaattcttt ggtccggttg  
60481 aactactctga tttaaaccaa gtttaagtgag atctccaca tatcgaaac gatcttgatt  
60541 caggtaacca aaagctaac gttaattcag atataacca aacgaaggga acagagagtt  
60601 tacacagcta cgggtctgtt tttgtgaca agtgtttgat acaaatttaa gacgaacta  
60661 aaatgggatt tagaaccctt gtacaaactc aggcattgta accttaagtt ttacatttct  
60721 tacattaaact agattggaac agtgtgctct ctcaacttta accataagct tgtattttgt  
60781 tgcctgccaac cggattTAggc gaggttagct tgctgtccct tcagtttgct gcgcgggaag  
60841 tgcgaactctg caatcaaaagg ctccgttccc ctctgtcttc gatcaaatcc acgtacacat  
60901 acgtacccta cataatatca aaagataagt tatgtttcag aacaagaaga aactgcttaa  
60961 tacaaaaatgt accctttccaa aagcaagcct gtatctcttc agttgataaa cctgagaaaa  
61021 atagagctcaa agtgggttaga acaactttct tttataataa caatcgcatc acaatccaat  
61081 aaagaaaaac ttataccttt gaatatcgta ggaacagagg taccaaaaa acgcgttctt  
61141 cagggtaatt cccatatcaa ttcccttggg acattgttgg ggtttaggct ggaatcgatg  
61201 tcgcgaacaa cctgtatcaa tagaatacat cacaagtttc aatgcaata attcaaatga  
61261 aagagttgga gttattggag ttcaagtctt accctatta ctgaaagta cgttccattt  
61321 agaggaaaaac tacccctcat cgtgttctca caaggaatct gtacaattta caacatatta  
61381 actctgtagaa aacataagtg tagtaagccg cataaggaga ttgatgcaac tacttaccaa  
61441 aattgtccct ctcaaatatt gagatctagt ctctctgatg ctgttgcagg agaacaagt  
61501 ctctcgtcca caaagcctac calttgcttg gaatatgcac gtaactaacg acggttgaa  
61561 agaatacagc gtctcacctg ttgaataaca catogattaa agataccgat ttgattccat  
61621 gattaaaaaga tatgcaaatc attaaattac ctgggttcca tatagcaagc aaataagaac  
61681 atggatcctc aggttctctc ttttccaaact gccacaagaa atcacaaaaa gctagtccga  
61741 tttaacataa tagacagcac ctatacaggg atgtgtcctt atccagttag ctacatacc  
61801 tgagctagaa gaggatgctc gtctgggaat tegttaactgc aagatacggg aaaaagaaca  
61861 agttatgcca tagcctgtaa ttattgggaa gttgtctgc tttccaaact acgagttcat  
61921 gcttggtcaa tcaacttaatt attctactct gtccaagctt taataatttt gaaaaatgtg  
61981 ttcttgattt catttttaac ctaagaacga agaaaaacag agaaaaatgg attcttcaac  
62041 tcggtgtctt gtcottaact ggctgatatt cttagaccta ggcattggaa gagaagcagt  
62101 ttcagcagta agtgcaacta aagcgtctga catgtttccg tcttgagttt cctgtgtgtg  
62161 ttccattatc ttcttcaagt tactggtaaa tgcattccatg tttagcctga cgttaggaat  
62221 ttcttctgga tctccaaaaa acgctctctc tatgtcagct attgataact ctgogggttc  
62281 tgggtccggt gaagcaggtc ctctgatgat tggttcaca catgtgacct ttttggctgg  
62341 ttctgagctg tgtactactt cagacccttg ctctctctgg aatggctctg gcaggtgtgag  
62401 aggcacaaagg tttttatcag gtgtcccat accctttctc gtaacttgta aagcaagcct  
62461 tgcactgtaa aacaatgtga tgttactaaa ttttctgtaa tgatgatcaa gagcttcgtt  
62521 tagatacaga ccaattctca tttaactggg ttatatatta acaaggactt tctcataga

62581 gtcatagtgg tactaaagggt ttaagagaac atgtgttagc accttgcaaa cgcactggca  
62641 aaatgtctgc attctccttt catcggacat gcattgcaat taggtttgct ottttgtgcaa  
62701 aagacctgat acaataatca agcagattac aaacctcacc atgtgagctg atttgaacat  
62761 acgtatataat gtatttcttt aatacatacc ttcccaaaag taatcatctg gtatgcaac  
5 62821 tcatacctgt gagataatag ggtattaaac taatgaataa gtgtattaga ctgaggcatg  
62881 aaaaaaaaaa agtttagtgat aaacatcatt cttacaatgt ttttggctg agtttcgaga  
62941 gacggggcca aagatacttt tgaatagatt caagcatagg atacctagac aaaccaaac  
63001 tcagatgtat taagtaacaa attacaattt ccaagtagga ccattttgaa aagtgtctac  
63061 atttccagaa gatgcaactg aagtgaactt gggagcggct gaaggggcacc ccatacaagt  
10 63121 ctgacggcta tgogcccaac atttgatca accgttcaat aaattaagtt catgcatcat  
63181 ataattcact ttttataggg acagaacaaa aagtttgatc ctgtcttact ggaaggcaca  
63241 gatggtagag ttttagaagc cgcacacact ccacacttt cagtcccaat ccgttaaaagc  
63301 tcagaagata ttctctgcag ggtttgttaa tatacgagag tacataatc attattaaat  
63361 cactaaaaact gccaaagtag taactcttgt ataggttaat aaagaagaaa taaaatgctt  
63421 cgtcttccaa acttactttg ctttatctgg tggaaacatct ctcaacact ccatacaagt  
15 63481 acttccatgg tcatttaccg gtccgatcaag gaagccctgc atgattttca tgttcagagt  
63541 caaatactta aaatgaatgt tatcacgaaa tttagccact aaatttttacc ctgtatacgt  
63601 tctgcgaagt tatggttcat ccgcgcactc ttgattgttt cagcaacttc ctttaacatg  
63661 gctgctcgta ttgccttcca atccacgggt tccattgtac ttctgtttt ttctctaat  
20 63721 cctgctctag cttgggcttc tcttctttaa caatccagct caaacgcttt tttttctctc  
63781 ttcaaaaact ttctcctttt ttctttaaag taggttttct gacacactcaa aacatcccca  
63841 tcttcttgag agcaacaacc aggttcacta ctatcaacag aggtattttg gggctcttcc  
63901 attgactgga aaaccttaat ttctgagcta caatcacccg gagacataat tgggtatcat  
63961 tgattggaat cttctagaga gacttgcacc cccgttagga gcttcaaaa ggaagtacga  
25 64021 tactctcctt ctaagtctat cttctgagct gatcctgtcc tctctgtatt ttgaggagca  
64081 tcagacacca tcgaatgttg acatgtaagt gcagatcttt cagatggaaa caggttccag  
64141 acacgaacct tctcatcgtt cttatcaacc cctgttagga agccttttgc atctgaatca  
64201 acatactcct ttgaatcgtt ggttttgcca actgattcat gggctgagat ggcactctca  
64261 ctactgcttc tggagggttc attgctaggt acataactct tctctcactc aggcctgtgta  
30 64321 tttttcaaaag taacagaact gtgattgtga tgggtgggc ttgacactgt ttctctgag  
64381 tccaagtacg ttatttgaat agaaggcatc gagctgtgtc cagcgtcaaa gttactgtcc  
64441 ggtacaaaag ggacagggaa ctggggaagcc aacgacatga aagcogaact caaaggagta  
64501 aaaaacatca aagcaagtta gttttgtgac tttttgctgt cttggattta gtttgacata  
64561 gaattatgta agagcttgta ccttgagaga tggctgaaa cattttgag gagaaatcat  
35 64621 ccaacaacag aatccacgac ggaacctctc caaggcgttaa aacgtgactc cccgtgtgaa  
64681 aaccaaaagc cataacaaga agcagtagct gagacatact aattgaaacc atgtgggttag  
64741 aacagaacaa cataaaaaga caagtgtggt gtataacctt gtacaaaggt catccttgca  
64801 ataatagagt cagctgtctc tcgaaaacaa ttactgttct cctccaccca tttgccttcc  
40 64861 tgctgtctg atcctgcaac accttgccta ttaattattc ccaatagcag ttctccact  
64921 ctgtctgtct catcgtctag atcaaccttt ggtctgggg gtgggttttt aacaggagtt  
64981 acaggcaca ttgtctcagc gccacaacca aagagtacaa tctggctatt catctgtgaa  
65041 ggaacgagag cagtttcaga atgctccctg ttgatgtcta atagacgcaa tagctcactg  
65101 attgtttcga tcgagttacg tcgtttccaa agttctcatg gagaagagcc tcaggaatcc  
45 65161 aaacatcatc attatcaaga aatagttctg atttaacaga ttcaaaaaaa caaagaaata  
65221 tagttctgta tctattcatt accagtaaat gaagggtgaa aacggcaga attcgggaaga  
65281 agatttttct gtttggttg ttattttcca gacttggttg gctttttttt gctcacagtc  
65341 acccgctgat caacaagaat cttattattt tgcgtggctc tgctacaata tctgagggtt  
65401 aaatgctcag gaacttccac aagtttattc atbgaagaca aattccoggaa tcgagcatgg  
50 65461 cctgtgtttt tcttctgtat ogcggtgtct atcttccog agagaacatt tagctcacag  
65521 ttgtccttcc cttgaagaat accttctcgt cttcttccaa cggtaggac tctattagat  
65581 ctgagctgag atgtggaaat gtgacacgta cgttgaaggg cagaacacata attgtgctgt

65641 tgtccagaga atactttgct gcaaatggca tcatatctca tccctccctg ttgcaagttt  
65701 ggggaataca accaatgcat ttcttggtag tccattgtct catcatactg aatccctgat  
65761 agattcatgt cgcaacaagt tggctcttgg ttctctcttg cagcttgogc catcgaaata  
5 65821 tgactattct ctgagttact gccatttttc ttggttgtaa ctccctttct ttactcttgg  
65881 ctgcgcttcc tcttgggctg gctaggagcc gaaagaagcc aatcttttaa ctcttgatcc  
65941 cagaatacta actgcttctc ttgagaagct tgatccatct cactgcttcc tcttaagtca  
66001 cgttggctct ggttttctcc attttcggct tcaaaatcca agactcgtct acagagctcc  
66061 ttaggacgag ttgaagtttc aacagctgct gatgattcaa ccggagtagc gtcttgatcc  
10 66121 ttactgactt caacctctct ccgcacatat ttctctcttg gtgtttgtct ttcttgacca  
66181 ttggtgacaa cagaacttct cggagctcgt ggtttaggct cctcttggg tttagcttct  
66241 ctaogaacct ttggcogagt cttcttctcc ttaggttttt caggagttct gaggactcgt  
66301 tcagcaacat tgttactcac tgagctcaa ctctccactt ctcagtagc aagcaaatcc  
66361 acctctgtgt ttctacactt gagaatcaca tcttctccag tccaatcaa acagaatcaa  
66421 aatttgacaa agogatttca ttctctatga gaccagaatc aaaaatccct cttaacttga  
15 66481 ggtattggag tctgaccaga gaatatgagg gatgcagtgt tagctagagc aagcaaatcc  
66541 ccaaaagaca agtgatcaag accactcata tctctgttcc caacaaatct cctgcatgca  
66601 tcaatacctt acttaaccaa ttaccatcca ctactcttgg aaattttcca actttagAAC  
66661 aaaaagcac aaacctttcc tccaattgac tgcctatgata ttgactctcc accgtgtatg  
20 66721 ggcagatcgg tgaaaaatggc ttactgggtg tctgaggaaat ccactggagg ttgtgaaagc  
66781 tgcttctctc tctctctctg ttctcCAItt ctgactctat tttaactttt cctcactctt  
66841 acttaaatca gaaccatttg agaaaaagct tggaaacttc ttatttttcc actcgcaaaa  
66901 gttcaataat ttcttcaata aaagagatcca ccaatttttt ttaaaaaatc cgaattttata  
66961 aaatgatcag atccactttt ttctggggtt tttagaagag agagatctcc ggaagtcatt  
25 67021 gattttgggt gagtggcgag atgaaacgat aatccgttct ttagggtgaa gagagacttt  
67081 ttgattctac aacaaaatgt aaaaaaaagt aagaaaaaaa caaaatctat taccagttag  
67141 atcaatgggt atgggtggtga tggagagagt tagttcgggt gtatgctatga gaggataaga  
67201 tcaactgatgc ttctgttctt ctcttggaaat cgatgaagt aaagagtaat atagaaaatga  
67261 ctttttttggc ctacgtgata aagaagagga tataacatgt gttgttctgt gtttcaactat  
30 67321 ttcttcaaac cgtttgttta ttaggggoga aagttcgttt ggttggcggg aaaaagttta  
67381 cggaatttta tttaaaaaat aatgattctt ttctacaaaa tctcttagac tttaggaaag  
67441 atgattttaa aagttaataa tattgtcgtt gttatcgtca tgcctatcat cgtcttttct  
67501 gttatctttt tctctttaa atttctgtat ttctctgtt taogtaacta tttaaaatta  
67561 tatgaactaa ctaattttat aattaataga aattataaaa taactttaa ttgtctttag  
35 67621 atataaaaa attagaactt tatttataaa ttatcatcaa atattatgatt taacaaata  
67681 acatgtttatg taatccacgt ttataatttt gatcaataat atattatttt ccttaattttt  
67741 acgtaacttc ataaatttacc acgttttctg ttacatatgc agaagttaa ttgattcgtt  
67801 tagaattatt atttccact gatattggag ctatgttagt agagtattga ttaggctagt  
67861 gtcocaaacga gtcttctgtt ttgtatcatt ccaaatgttt tagtctagta ttgtaggagt  
40 67921 caaaatactg caccatattg gtgaaactgt gaattgtgtg gaaaaaaga gtaattagt  
67981 tgtaacctt tgatttctct tcatgcaaga acccttcaaa gagagctaca tgagaaatga  
68041 gttatgttaa tcaatttatt catggacttg gttggaatct tagtgaatct ttgttgcaaa  
68101 tcttaacaac ttgttggtatt ggttatgagc ctatgactta tgacttatga gtgagtcaat  
68161 ggttggtcata acctaattat tgggttatga gcaaaagaa ttggaatttg taaaaaaaa  
68221 aaaaaaaact agagctttt ttgtgtggac atatctatcc tagaaactga gacgaataat  
45 68281 agtggtataaa aagtgggaa cggattatcc gaattgttaa aactattatt gaaaacaata  
68341 caactaaata tgggtacaaa gtaaaogaat togtatagct aaacctaat caaattacga  
68401 agctaatacca tacttggtac ctaaacgctt ttacttttac ttacggttcc tttttcaaaa  
68461 aagtttttacc aaatttgggt ttgtcttatg aagattatgg cagaagagac tgatcaaaag  
50 68521 tgaagtccata attcgttata atccattcaa gtttacttta aacaatgaaa ctgacatga  
68581 aagtgaattc aaagaccaaa tcaaaagaaa attaaactga tttagttgta attttggtat  
68641 tgaattaaac tataaataga aataaccaa catataacca caaaagaaga ctatttatat

68701 aaatatatga gttggaagtc attttggac tattatataa gatctaatta tcacacgaag  
 68761 tgtggatgta tggtagcag agttgtgttc agagagttcg ataaagccat cactccaaac  
 68821 atacaaaata tccatacatt gatccacca tataacoggc tgtgtgcgaa gcaagtgaa

# 5 SEQ ID NO:11

ARABIDOPSIS THALIANA DMT3

>DMT3 (1DMT3) ;

MEVEGEVREKEARVKGRQPETEVHLHGLPQEQSIFNNMQHNHQPDSDRRRLSLENLPGLYN  
 MSCTQLLLALANATVATGSSSIGASSSSLSSQHPTDSWINSWKMDSNPWTLSKMQKQQYDVS  
 10 TPQKFLCDLNLTPPEELVSTSTQRTEPESQITLKTGPKSLSETDHEPHDRIKKSVLGTGS  
 PAAVKKRKIARNDKESQLETPTLKRKKIRPKVVREGTKKASSKAGIKKSSIAATATKTS  
 EESNYVRPKRLTRRSIRFDFDLQEEDEFCGIDFTSAGHVEGSSGEENLTDITLGMFGHV  
 PKGRRGQRRSNGFKKTDNDCLSSMLSLVNTGPGSGFMESEEDRPSDSQISLGRQRSIMATR  
 PRNFRSLKKLLQRIIPSKRDRKGKCLPRGLPKLTVASKLQLKVFRKKRSQRNRVASQFNA  
 15 RILDQWRRQNPTGTSLADIWERSLTIDAITKLFEBLDINKEGLCLPHNRETALILYKKS  
 YEEQKAIKYSKKQKPKVQLDPETSRVWKLMSIDCDGVDGSDDEKRKWWEEERNMFHG  
 RANSFIARMRVVQGNRTFSPWKGSVVDSVVGVLFTQNVADHSSSSAYMDLAAEFPVEWNF  
 NKGSCHEEWGSSVTQETILNLDPRGTGVSTPRIRNPTRVIIIEIDDDENDIDAVCSQESSK  
 TSDSSITSADQSKTMLLDPFNTVLMNEQVDSQMVKGKGHIPTDLDLNDLSQGISMVSSAS  
 20 THCELNLNEVPPEVELCSHQQDPESTIQTDQQESTRTEDVKNRKKPTTSKPKKSKES  
 AKSTQKKSVDWDSLRKEAESGGRKRERTERTMDTVWDALRCTDVHKIANIIIKRGMNM  
 LAERIKAFNLRLVKKHGSIDLEWLRDVPDPKAKKEYLLSINGLGLKSVECVRLSLHQIAF  
 PVDTNVGRIAVRLGWVPLQPLPDELQMHLELYPVLESVQKYLWPRCLKLDQKTLVELHY  
 HMITFGKVFCTKVKPCNACPMKAEBCRHYSSARASARLALPEPEESDRTSVMIHERRSKR  
 25 KPVVVNFRPSFLYQEKEQEAEQRSQNCPIIEEPASPEPEYIEHDIEDYPRDKNNVGTSE  
 DPWENKDVPTIILNKEAGTSHDLVVNKEAGTSHDLVVLSTYAAAIIPRKLKIKEKLRT  
 HHVFELPDHHSILEGFERRAEDIVPYLLAIWTPGETVNSIQPPKQRCALFESNNTLCNE  
 NKFCQCNKTREBSQTVRGITILPCRTAMRGGFPLNGTYFQTNEVFADHSSINPIDVPT  
 ELIIDLKRRVAYLGLSSVSSICKGLSVEAIKYNFQEGYVVCVRGFDRENKPKSLVKRLHCS  
 30 HVAIRTKETEE

SEQ ID NO:12

>DMT3(1DMT3) novel 375 amino acid amino terminus;

MEVEGEVREKEARVKGRQPETEVHLHGLPQEQSIFNNMQHNHQPDSDRRRLSLENLPGLYN  
 35 MSCTQLLLALANATVATGSSSIGASSSSLSSQHPTDSWINSWKMDSNPWTLSKMQKQQYDVS  
 TPQKFLCDLNLTPPEELVSTSTQRTEPESQITLKTGPKSLSETDHEPHDRIKKSVLGTGS



PAAVKKRKRIARNDKESQLETPTLKRKKIRPKVVREGTKKASSKAGIKKSSIATATKTS  
 EESNYVRPKRLTRRSIRDFDLQEDEEFCEGIDFTSAGHVEGSSGEENLDTTTLGFMGHV  
 PKGRRGQRRSNGFKKTDNDCLSSMLSLVNTGPGSFMESEEDRPSDSQISLGRQRSIMATR  
 PRNFRSLKLLQRII

5

SEQ ID NO:13

>DMT3(1DMT3) nucleotide sequence from BAC T22K18 (gi 12408726);

53341 aatcaagtac taatgcagat ttaagggggg tgtattgacg gcgttaaac ggtttctcaa  
 53401 cggaaatcgta cgtagtacac cgtgatttta ttgtttaacc cggattggtc atgctgttct  
 53461 tcttttccac ttgcgcggac cactcaatga cactctcttc ttttgtagca ttggcccgac  
 53521 accagaatgc agcatttaat ctctcaaat accattttgc tctcactctc tttaccctct  
 53581 ttgtgtatttt gtgtcttttt tctttctatt tctgtgtgaaa aaggatctct tcttaaatcg  
 53641 tattatttct tcgatatact accttttatt tgttttctat ttttgtagg ttacatcttt  
 53701 tttataaaga aaatatgagc taacacagca ttagtgtgtg taaccaaga atgggaaaaa  
 53761 agttataaga gagataataa gattctctta cagagactca cttaagttaa aaagggaaga  
 53821 cgaagtgtgt ccttaaggg aaaaaaagt cactgaactt catatacac ttaataactg  
 53881 actgtgtaac tcaatagatc gtgcagtaac attcagctgt attagtaaga aggaatttat  
 53941 ttgctaagta aactcaagcc tcttttttct ctttttttct ttttagtaa aaattaggct  
 54001 agtggttttt ttgactcagc aacactctgc ttaaaattag gagtaatttg acctattctc  
 54061 actgattttct aagtgaattc tgttgggggt aaagaagcaa ctgattgaat tagtggaaaa  
 54121 tctgttctct tctttacgca tagttcaact tggacactca gtctcaatgc ttctacgttt  
 54181 cactgacgaa caacatatat tcatcagttt gtgactcgtc catctgtgat aagtgtgcat  
 54241 tcaagtgaac ctgcacacac ttgtgtgcaat tatttggcgc tcaatctat tagtggaaaa  
 54301 tttacaatct aattgttcta ttatttcatt tacttgtcat caatttatta tatttgtagc  
 54361 caatgaactg tgtaattaaa gaacaaaaat aaataaatat ctggaatttt gtaacagcta  
 54421 ctagaagctg atttcttatt aattgtatca ctaaaattat attaaaaagc aaactcaaga  
 54481 atgataatta tatatttaaa aaatttctgt tgtcacattt cttttaaact acaattatga  
 54541 atactcaaaa ctacttcagc catatcttaa aatttgaatt caaaactctc ttactcttct  
 54601 tttagggtct taattaacag tcaactaaaa tagtcaaggt ttggaatttt atgaaaaaag  
 54661 ataagatgat aattaatgga tacgcctcgt acaaaattct tgtaaaatgt agataatata  
 54721 cattttgttaa atatgacacg tgtttatttt ttttttaaat atgatcaaaa tatatttttaa  
 54781 ctactctagat ggtagtgatg tctccaattt tgaataacaa gtcaatttgtt attagagaattc  
 54841 tcaataataa aagaaggga ttaaaattgc aaagaaaaag tgaaaaacaa aggatttgtta  
 54901 tttttggaaa aattaaggac tggatttgca aaaaagaaa agtaacttca tgtatattga  
 54961 ctctcttata gtctctataa actattatct caaattttgt ctggactctg aaactcaaga  
 55021 gaactgaact tggcttactt ggcttcatct tttctctctc ggtaactctc cctgcaactt  
 55081 caagctttca ttttcaata aatgtaatca aatctgttat tttcactcaa gaactaattg  
 55141 agttctctat ccttttcaat tgaaattgac aataaatga aaagattttg aggaggtttc  
 55201 aactcaaca accgaatcac ttcttctctc aaatatttgt tcttctagtg gccaaagatc  
 55261 acaatcaatt ttgtatctct ccacaggtaa attaatgtgt attgaacaga gaagaggaca  
 55321 agtgcactgt gtctcaaaaga aATGgaagt gaaggtgaag tgagagaga agaagctagg  
 55381 gttaaaggga gacaaccaga gacagaagtt ctacatggct tgccaacaga acagtcaata  
 55441 tttataaaca tgcaacacaa ccatcagcct gactcagaca ggttttgtga ctcaaccgaa  
 55501 ttactctgt tctctctccg gaatttccat atttctgtgt gattctgttt tgttaaaatt  
 55561 tgcaaaaggga agaaaataaa tcaaacattt ttactctctt caaaacatga gtaaatgcaa  
 55621 aaactgagat atgtaaacac acagcaattt ttgatgaac tgggtttggc tgtgtgactc  
 55681 ttgtgtctat gcaattacgt tttagttatt tctacttta taaggagaga tgttaactga  
 55741 aactgttatt gatcatacag gaggaggctt agtcttgaaa acttactcgg actatacaac  
 55801 atgtcttgta cacaactctt ggctctggcc aatgccacag tcgccacagg ttcactcaatt

CGA40743.042301

5

55861 ggtgcatcat catcatcggt aagctctcag catccaacgg attcttggat taatagctgg  
 55921 aagatggact ctaatcogtg gactttgagt aaaatgcaaa aacaacaatg tgagtaaaat  
 55981 tgttctcctga atttgtagga tcttttaaga gaaagtaagc gttttatgtg atgattagtc  
 56041 agactgaaat cgattatctc ataataagtt ctacgtgac tctcaatca tgaattttat  
 56101 gtttaacctga tatcaacttc ttgcttgggt gaaccacaga tgatgtgtca actcgcgaga  
 56161 agttctcttgg tgaccttaat cttaacacct aagatgtggg gacacccagt acgcacagaa  
 56221 cagaacactga gtctcctcaa ataactttaa agacaccagg aaaaagctcg tctgaaactg  
 56281 atcatagacc tcacgacogt atcaagaagt ctgctcttgg aactggatct cctgcagcag

10

56341 taagaaaaag aagatagca agaaatgatg agaaatctca gctggaaaca ccaacactaa  
 56401 agagaaaaaa gatcaggcca aaggttgtcc gtgaaggcaa acaaaaaaaa gcatcatctca  
 56461 aagcagggat taaaaaatcc tctattgtcg ctactgtctac taaaactctt gaagsgagag  
 56521 attatgttcg gccaaaaaga ttaacgagaa gatctatacg attcogactt gacctccaag  
 56581 aagaagatga ggaattttgt ggaatcgatt tcaatcagc aggtcagcta cgaaggttctt  
 56641 caggtgaaga aaactaacc gatacaacac tgggaatggt tggtaacgtc ccaaggsgaa  
 56701 gaagagggca aagaagatcc aatggcttta aaaaaaccca caatgatgtg ctacgtttcta  
 56761 tgtgtctctc tgtcaatacc ggaccaggaa gtttctcaga atcagagaaa gatcgtcoga

15

56821 gtgattcaca aatttctctg ggaagacaga gatccattat ggcaaccaga ccogctaatc  
 56881 tccgatcggt aagaagaact ttacaagga ttataccaag caaacgtgat agaaaaggat  
 56941 gtaagcttcc tctgtgactt ccgaagctta cctgcgcatc caaggtgcga ctaaaagtgt  
 57001 ttagaagaag cgggagtcga agaaacogtg tagcaagcca gttcaatgca aggatattgg  
 57061 attctgcagt gcgaogccaa aatccaacag gtgataaaca cacaagcaac ttcatctcat  
 57121 aatatttttc tttagatttct atcttttgaa ttaatactag ttttacaaaa tgcaggtaca  
 57181 tctgtagctg atatatggga aagaagtgtg actattgatg ctatacactaa gtgtgttgaa  
 57241 gaattagaca tcaacaaaga ggggtctttg cttccacata atcagagaac tgcacttatt

25

57301 ctatacaaaa agtctgatga agagcaaaag gcaatagtag agtatagcaa gaagcagaaa  
 57361 ccgaagatga aattggatcc tgaaaacgag cagagtgtga aactcttaat gtaacgtatc  
 57421 gactgtgaog gtgttgatgg atcagatgag gaataaacgta aattgggtgga agaggagagg  
 57481 aacatgttcc atggacgtgc aaactcgttc attgcogaga tgcgtgtgtt ccaaggtatt  
 57541 attttattgt tttagttatga cattgttgtg tggctttata ccttagatct tcttttcttt  
 57601 cttttttgta tccaagcaaa catggtctta aatcaagctt atcactgcag gcaatagaac  
 57661 tttctcaact tggaaagggt cagtagtgga ttcagtagtg ggagttttcc taaccagaaa  
 57721 tctgcgacag cattcatcaa ggtatatgca ttcaagagat ttctaataag tagaagatat

30

57781 atgcaacaga gtggtttaga aattataact tgttctactt tgcagttctg catatatgga  
 57841 ttttagctgt gaatttctct tgcagtgaaa cttcaacaag gatcatgtct atgaagagtg  
 57901 gggagattca gtaactcaag aaacaatac gaatttggat ccaagaactg gagtttcaac  
 57961 tccaagaatt cgaatccaa ctcgctcat catagaggag attgatgatg atgagaacga  
 58021 catgatgtct gtttgtgtc aggaatcttc taaaacaagt gacagttcca taactctcag  
 58081 agccaactca aaaacagatgc tgcctggtac atttaacaca gttttgatga acgagcaagt  
 58141 tgattcccaa atggtaaaaa gcaaaaggta tataccatac acggatgatc ttaatgactt  
 58201 gtccacaggg atttcogagg tctcatctgc tctactactc tgtgagttga acctaaatga  
 58261 agtaccacct gaagtagagt tgtgcagcca tcaacaagac ccggagagta ccatctagac

35

58321 acaagaccag caagagagca caagaacgga ggaatgtgaag aagaatagga aaaaaccaac  
 58381 tacctccaaa ccaagaaaaa agtcaaaagg atcagcaaaag agcacgcgaa agaaaagcgt  
 58441 tgaatggagt agtttagaaa aggaagcaga aagtggtggc cgaaagagag agagaacaga  
 58501 aagaacaatg gacacagtgt attgggagtc acttcogagt acagaogtac acaagatcgc  
 58561 taataataac atcaaacagg ggaatgaacaa catgcttgcc gaaagaatca aggtttgact  
 58621 aactcacagt ctatatatac cttattata catctcaaca agtggaattt ttttgactct  
 58681 ggaattttaga caggctcttc taacacagact agttaaaaa catggaagca ttagacttaga  
 58741 gtggctaaaga gatgtcttc ctgataaagg caagtgaaga aattatttca aaactttgag  
 58801 attatatgta gccctcgggt aaagaatata tctcagtaaa tggactcgat agtaattgtg  
 58861 atacatatata atgagagata cttgatagtg actactaagt gttgcaggga gtactacta

40

45

50

58921 agcataaacg gatttaggatt gaagagtgtg gagtgtgtta gacttttgtc actacatcag  
58981 attgcatcc ctgtaagtca atgaaggata ctgaatactc agaccctaataa gaatgtggaa  
59041 cagatacatt aatagttacg tatttttaca aatgcagggt gacacgaagt tgggaagcat  
59101 agctgtaaga ctaggatggg ttcccttaca gccattgccc gacgagctgc aaatgcattc  
59161 ttttagagttg taagaaaaaa aaattaaaga tcaattcttca atcatgaaag ggaacatgag  
59221 aaattttacag tagttccctt taattctatt caggtaccoca gtcttagagt cagttcaaaa  
59281 gtacctctgg ccacgcctct gcaagcttga ccaaaaaaac ttgtaagtaa attacattag  
59341 catcaacocat tactctagac ccttaaacct ctctaactaa ctctaactgt atcatacaat  
59401 cttaggtacg agctgcatta ccaactgata acatttggaa aggtacctca aacaatttcc  
59461 aagtgtttgt ggaatgaaaa catcttaaaag tggcttttcc tattttgcag gtcttttgca  
59521 caaaagttaa acccaattgc aatgcattgc caatgaaggc ggagtgtcga cattactcta  
59581 gtgcacgtgc aaggttaaac cccacaaaat tctttgttat tgccattaac atgaaaaaaa  
59641 aaacactago ttaaaagaaa agagatctgc tcaaaaatag cattttaatg gttgtatggt  
59701 ctaaaatgctt gtgttatatc gcagcgcacg gcttgcctta ccagaaaccag agggagtgta  
59761 caagacaagt gtaatgatcc attgagaggag atctaaacgc aagcctgttg tggttaattt  
59821 tgaccatcc ttattttctt atcaagaaaa agagcaagaa gcacaaagat cccaaaaactg  
59881 tgaaccaatc attgaggaac cagcatcacc agaaccagag tatatagaac atgatattga  
59941 agactatcct ogggacaaaa acaacgttgg acatcagag gatccttggg aaaataagga  
60001 cgttaattcct accatcatcc tcaacaagga agctgggtca tccatgatt tgggtgtcaa  
60061 caaggaagct ggtacgtcac atgatttgggt ggtaactaagc acatattgca cagcaatacc  
60121 tagacgtaaa ctcaagatca aggaaaaagt acgcacagag caccactgtg gagttgccac  
60181 ttccaatttt ttctctctatt ataccctaaa ccgtaaaaatt tgagacttcc ctacgacttt  
60241 atctcactact aattctcttt tacagatttg agctccctga tcaacttcc atctctagaag  
60301 ggggttagtaa ctcttgcaaa atgattttag aagaattttt ctacttattc ccgctctaaa  
60361 aactgtttga ttatcttttt ttacagtttg agaggcgaga agctgagtag atagtccctt  
60421 acttgttagc catttggacg ccaggtaaga agaaataggg acacaataaa atctgattat  
60481 gatattttctt ttcaagaata ccgctatat tttaacaggt ttcatcctta gatctattatg  
60541 actaatgtct aacaagtgat tgtaatat ttccatacca ggtgaaacgg tgaaattccat  
60601 tcaaccgcga aacaaagat gtgctttatt tgaaggaat aatacattat gcaacgcaaaa  
60661 caaatgtttt caatgcaaca agacacggga agaggaatca cagactgtac gaggaaactat  
60721 attggttaaga ttctggtgga caattttcaa gagaatatct ctaagttaga atataaggaa  
60781 ggtataaaaa tgactaattt gtttgttaac agataccttg cagaacagca atgagaggtg  
60841 gattcccttt gaatggcaca tacttccaaa ctaatgaggt aattttccca aaaaatgaatt  
60901 taacttaaac aaatgatcaa aagcaacatt ctctgcaagg ctgcatttgg actataactg  
60961 tgcaggtttt tgcgtaccat gactctagca taaaacctat cgaagctcca acgaactcga  
61021 tatgggattc aaaaagaaga gtccgatact taggatactc tgatctcgt atttgttaag  
61081 gtaaattttc aaaaacaaac tgcgtagtta tgcatgtgtt tggatatata aatccaaagt  
61141 ctgtgtccaa tatgtttttc tcaatttttt aggtttatca gtggaaagca taaaatacaa  
61201 ttccocaggaa ggtatgctaa tatgtcttac actgaaaaaa ccttttagtat caaacattga  
61261 attcatgaaa agaacaacaa atagatataa aatcagtcac gatgtttttt ctttggogat  
61321 tgaagatgtt gataggaaag tatagaagat atagcttaag ttggtttaata ctgtttttat  
61381 agagctttga ggtgggtttt gactagcatt gtaatatata tgcaggatatt gctctgttaa  
61441 ggggatttga cagggagaat cgtgaagcaa agagtctagt gaaaagactg catgtgtctc  
61501 acgtagcaat cagaactaaa gagaagacag aggaatgaaa ccttcagat tgctattaaa  
61561 tgttagacat atttgaatca ttgggtttagg gtttaccatc caaaggtcat agaggatctt  
61621 atttttccat taacttttaa attcatgcaa ctcttttttg gtgtttcttt ttgttctctg  
61681 ccatagtttt gggcaatgga tggagtgtct ttgcaaaact aggttttttg tactcattaa  
61741 cagaaatttg cagcaactaa tcatctttcc tattatctat caaagctctc agtgtttctc  
61801 cataacttga tgagatttag tcaactctca gctaatcag tctgttccca atttccatca  
61861 gatttggtaa aggaacaact gcaattgcta agtacaacat gatccagatt tcaaacaggt  
61921 tccaggttta atccaaatca tcaatttcaa tcaagagca aactagaatt caaaacatat

61981 aatctctgat tcagattcaa gaaagacaaa gcatgagaca tcattctgca agttaaccaa  
62041 ttccggttat tctcgaatcc tactgaatta agcatcaatc atctaaagga acttctaatag

# 5 SEQ ID NO:14

Arabidopsis thaliana DMT4

>DMT4 (1DMT4) ;

MEFSIDRDNLLMVVPETRIKTKQFEKVYVRRKSIKLPQNSMVHNTLIKMARQRIQKSMK  
ESVMNQHIKFNDFSYSLSVIYHPCCFVINNSQTTHKKKEKKNSKEKHGHIKHSSEHLQDDI  
10 SQRVTGKGRRRNSKGTTPKKLRFNRPRILEDGKKPRNPATTRLRTISNKRKKDIDSEDEV  
IPELATPTKESFPKRRKNEKIKRSVARTLNFKQEIIVLSCLFEDKICGPIFPRGKKRTTTR  
RRYDFLCFLLPMPVWKKQSRRSKRRKNMVRWARIASSSKLEETLPLIVSHPTINGQADA  
SLHIDDTLVRHVVSQTKKSANNVIEHLNRQITYQKDHGLSSSLADVPLHIEDTLIKSASS  
VLSEPIKKTDKIAKLIDMGRLLKINKVTTMIKADKKLVTAKNLDPETIKEDVLMVN  
15 DSPSRSYDDKETEAKWKKEBIFQTRIDLFINRMHRLQGNRKFKQWKGSVVDVSVGVFLT  
QNTTDYLSNNAFMSVAAKFPVDAREGLSYYIEBPQDAKSECIILSDESISKVEDHENTA  
KRKNEKTGIIIEBIVDWNLRMYTKESRPEMHMSVNSWSDVRLSGQNVLETTIKRGQ  
FRILSERILKFLNDEVNQNGNIDLEWLRNAPSHLVKRYLLEIEGIGLSAECVRLGLKH  
HAFPVDTNVGRIAVRLGLVPLEPLNGVQMHQLFEYPSMDSIQYLWPRCLKLPQETLYE  
20 LHYQMITFGKVCTKTIPNCNACPMKSECKYFASAYVSSKVLLESPEBKMHEBNTFMNAH  
SQDVAVDMTSNINLVBECVSSGCSQAICYKPLVEFPSSPRAEIPESDIEDVFPFMNLYQ  
SYASVPKIDFDLDALKKSVEDALVISGRMSSSDEEISKALVIPTPENACIPIKPPRKMKY  
YNRLRTEHVVVYVLPDNHELLHDFERRKLDPSFYLLAIWQGETSSSFVPPKKKCSSDGS  
KLCKIKNSYCWITREQNSNIFRGITILVFADHETSLNPIVFRRELCKGLEKRALYCGSTV  
25 TSIFKLLDTRRIELCFWTGFLCLRAFDRKQRPKELVRLHTPPDERGPNGPHIVVDEK  
BESPRVGLMVMPGFVWIGGSVIQNRVYVSGVKVLE

# SEQ ID NO:15

>DMT4 novel 372 amino acid NH2 terminus;

MEFSIDRDNLLMVVPETRIKTKQFEKVYVRRKSIKLPQNSMVHNTLIKMARQRIQKSMK  
ESVMNQHIKFNDFSYSLSVIYHPCCFVINNSQTTHKKKEKKNSKEKHGHIKHSSEHLQDDI  
30 SQRVTGKGRRRNSKGTTPKKLRFNRPRILEDGKKPRNPATTRLRTISNKRKKDIDSEDEV  
IPELATPTKESFPKRRKNEKIKRSVARTLNFKQEIIVLSCLFEDKICGPIFPRGKKRTTTR  
RRYDFLCFLLPMPVWKKQSRRSKRRKNMVRWARIASSSKLEETLPLIVSHPTINGQADA

SLHIDDTLVRRHVSKQTKKSANNVIEHLNRQITYQKDHGLSSLADVPLHIEDTLIKSASS  
VLSEPIKKTQD

SEQ ID NO:16

5 >DMT4 nucleotide sequence BAC F28A23 (gi 7228244);  
14881 gctatggatg tcaacagaga gaattacgaa ttgggtttac cgtatcattga gaaagccggc  
14941 gttgtcaca agatcgactt cagggaaggg cctgctcttc cgttcttga tgaatcgtt  
15001 gctgacgtaa gcattcttct ttctgacgta attaacaaaa aagatgatga agataatgaa  
15061 ataattaaaa actcatggcc taattagggt gatttaatat ctgtgataga attctgtgat  
10 15121 agocaaaattt gttctctttt tcatagaaga aagtgtggta actgatattt gtgtgtggtt  
15181 ggggtgcagga gaagaacctt ggaacatatt acttttatatt cgttgatgct gacaaaagca  
15241 actacatcaa ctaccacaag cgtttgatcg atcttctgaa aattggagga gtgatggctt  
15301 agocaaacac tctgtggaat ggttctgtcg tggctctctc tgatgcacca atgaggaagt  
15361 acgttctgta ctacagagac ttgtttcttg agcttaacaa ggctcttgct gctgaccttc  
15421 ggatcgagat ctgtatgctc cctgttggtg atggaatcac tatctcgact cggatcagtt  
15 15481 gatttgactc ctccctactc tgagtttgct cacagtgagt tactttocat ctctttatac  
15541 ctttcaatgc cattttccac aacctactaa atggaccttt ttatgtattt gtgttaagta  
15601 atactctcat tgccttggtt ttgctttctt ctgaacaaag aaataatgtg tactttactt  
15661 ttctttctgg tctcgtttct ttgtttttct ccatgataca acatctaaag aaattatttg  
20 15721 tgtcacagca acgttaagtc ataaaaattg ttgaacatat tgagaaaaag ttatcataga  
15781 ctttcaattg ttgaagtc atgttggtat ttgtcaattg atattagatt accaaataaa  
15841 tattagacag taagaaacga acaagtagg aagatgtagg tcaocgctct ttgaaaaatt  
15901 atcagataga attcataata cacagttagg tagtttctgt tgagagttaa aagggaaaaa  
15961 tatgtaatgt tgtgtgataa atacgtcaaa aattagttga tggagcaaat cgttaacaaa  
25 16021 aataactttt tgcattagtt ttgttggtat cctcataaat acgggttccc atattcaact  
16081 cgtagtttag ataattataa gcaacaaata aacacaaat actgaattta gaaatttttc  
16141 agaaaaataa tttagagatt tacattattt ctacaaact tagtgaatta ttctttaaac  
16201 gtatgttagt tatttattaa ctgaagtttc acatatttga tagaataaca tttaaatcaa  
16261 aaaaattgaa gtaaggttag aatgttctta taatacttta taactttttt aaaaggtaca  
30 16321 agocaaaatt atcgcaaatg taataataaa atcattgtaa aaactttaaa ctaattaaaa  
16381 gatctaacgc aatctaaaaa aagatttggt atcatcgccc atttatgttt tgataatac  
16441 aaaaactggt aataattaaa tttaattatc aatttcttaa ttagttagaa tctttgttaa  
16501 tgtaatcaac tcaaccttat tttaattat taaaattggt gttatatct cttaatac  
16561 tctaagatga tattttcttc catttatgaa aagaaaaata tgttaattaa gcaataaaaa  
35 16621 gaaggaaaaa ataattttaa taatatttaa tatatatata tggtttttag agttcgagtt  
16681 ctctcgatatt tacagtttct cttttttcca aagcagggtt tggatggta gtttttctgg  
16741 attaattttg tctcaaatc ttctctcttt ttattttttt ttgtgaaatt cttgttttta  
16801 attggtgtga catcgtttcc aaaatatttt caaatttgat tgcctttgaa gttttttttt  
16861 ttcttctatg ttttggaaat cattatacta gogttgttgt ttcttctctt gcaagagtaa  
40 16921 tggagtttcc aatagatoga gacaaaaatc ttctcatggt tgttccagag acacgtatca  
16981 aaacaaaaaa atttgaaaaa gtttatgtga gaagaaaatc tattaagctt ccaaaaaatt  
17041 cgttaatttt tccacatgaa atcaaaagat ggggtgaaga agagagtaag gagaaggaat  
17101 ttcttcacat aggttaacaa aatctctaata accttaatta ctccogttta gtaattctcc  
17161 ttttaactgt ttttttttta atgagagat gtgacaattt cataagaaa taagtgtttt  
45 17221 gacatacog atggtttttt gactaattat attttttgtt ttgaaagatt tccaagctaa  
17281 tttaattag catatttttg attttattga ttgaggaatt ttccagaatt tcgacattta  
17341 agtttttttt ttgtttttaa tatacttttg attogatgat aagagattgg gaaagcagac  
17401 taatgatgtt ttgtgtgac gttcattgat tagagatctc ttatatctat atttgtctac  
17461 aatatatcat gcatgtgttg atttgtttcg ttaattcaat tttttttttt tcatgttgac  
50 17521 agatgggtta caacacactt atcaaaatgg cgagacaaag aattcaaaag agcatgaaag

5

10

15

20

25

30

35

40

45

50

17581 aaagtgtgat gaatcagcac atcttcaagg taaataattt taaattcatt cttaaaaaag  
17641 tttagcttatt ggtaagtcca ttacaattta tatttaacca togtcacttt ttatttaacg  
17701 agtttgataa gcattttcaa aaacctgtct toactcgccg atgcagatgt ggttatgttc  
17761 acctcttgatt ttattgattg aggatttttt cagaatttcg attcactact gtctgtaata  
17821 tatcatccat gttgttttgt aatcagttaa ttccacttatt ttatttttaa cttttattgt  
17881 aacagataat tccaaaacca cccataaaaa aaggaggaag aagaattcaa aagaaaaagca  
17941 tggataaag cattctgaat cagaacactct tcaaggtaaa tacttttgaa ttcatctatt  
18001 aaaaaaacag tttatttgta agttcattac agtttatata tatttaaatt gttttatgta  
18061 agtatttttt gcacaatcga ctaactcata cccactcatt cattttatt ttattttatg  
18121 gtgaaagatg atatttgcga aogtgttacc ggaaggaagg ggaaggaagg ttcaaaaggg  
18181 acaccaaaaa aactgaggtt taataggcct cggatcttgg aagacggaaa gaaaccaaga  
18241 aatcccgcca ccactcgact gagaactata tccaacaaga ggaggaaaaa ggacatagac  
18301 agtgaagatg aagttatacc agagcttgca actccaacaa aggaagagctt tccaagaga  
18361 agaaagaacg agaagattaa gagatccgtg gctcggactt taaatttaa gcaagaattt  
18421 gttctgagtt gttctgagtt cgacaagatt tgtggacca tttttccaag agggaaaaag  
18481 aggaccacca cagcaogcag atatgatttc cttgttttt tacttcgat gctgtttgg  
18541 aaaaaacaat caagaaggtc taagcgtagg aaaaatatgg tcagatgggc tagaattgct  
18601 tcttcttcaa aactgctaga agaacttttg cctttaatg taagtctacc gactattatt  
18661 ggacaagcag atgcttcttt acacattgat ggtaactcag tttttttttt gttatttat  
18721 ctgtttacac aaaattgttt atgottatat ctaagctatc attgtgatt ttattttgca  
18781 gacacactcg tgagacatgt agtctcaaa caaacccaag aagtgctcaa caatgtcatt  
18841 gagcatttaa atcgacaat aacttatcag aaagatcag gtctctcact cctggcagat  
18901 gttcctttgc acattgaagg taacttagtc ttatttttgt tcttttttaa tactattgatt  
18961 aaaaagattg tgatataatt atttaataa ttittgttat attatactata tattttattg  
19021 ttgttacttt ttttttgtag atacactaat aaaaatggct agtctcgtao ttccagaacg  
19081 accactcaag aaaactaagg atattgctaa gttactcaa gatatgggaa gattaaagat  
19141 caataaaaaa gtaacaaoga tgatcaaagc tgacaagaaa cctgttacgg caaaggttaa  
19201 tcttgatcca gagaccatta aagagtgagg tgtcttaatg gtgaatgatt caccagaacg  
19261 atcatatgac gataaggaga cggaggccaa atggaaaaaa gaaagagaga tttttcaaac  
19321 ccgatatagat cttttcatta accggatgca tgcgtttcaa ggtacattat tgttattatc  
19381 attattgtta ttatgatcta ttataacttg tattctaaat tagottacat atatatata  
19441 ggaatccaag tataagttag tatgctaagt atatgatcat tttttgaaat tatgtttctc  
19501 tccatgatgt taaatgatt gtctgcagg caatagaaag tttaaacagt ggaaggctc  
19561 agttgttgac tcaagtgttg gaggtttttt gacacaaaat actaccgact atctttcaag  
19621 gtaaaatctt tgtttaaatt gttaaagaa ttgaaaaact aattctata atagatgatc  
19681 actttgattg tgagtttcta cagcaacogc tttatgagcg tggctgcaaa atttctgtt  
19741 gatgcaagag aaggtctatc atactatatt gaggaaacct aagatgtaa aagttctgaa  
19801 tgtatcattt tatctgatga tccaatatca aagtggaag atcatgagaa tactgcaaaa  
19861 aggaaaaaag agaaaaacog tattatagaa gatgagatag ttgactggaa caatcttaga  
19921 aggatgtaca cgaagaagag atctcgtccc gaaatgcata tggactctgt taattggagt  
19981 gaogctgagat tatctggcca aaatgttttg gaaaccaaca ttaaaaaacg tggacaattc  
20041 aggattcttt cagaagaat attggttaaga aaaaacaaac ttctaataag ctttgtgaat  
20101 aattttattca aatgatitaa gactaacact tttttttt tcttgtttt ctcaagaat  
20161 tcttcaoga tgaagttaac caaaatggaa atattgatct ggaatggctt cgaagtcttc  
20221 catcacattt agtgaagatg gtttatgttg gtttttatgt tctcatagat ctcatatta  
20281 gtaagcgatc ataaactctt tctattattt tatcaggaga tatctgttgg aaatcgaagg  
20341 gataggcgctg aaaaagtctg agtgcgtacg actgttaga ctttaaacat atcggtttcc  
20401 ggtatgaaaa tattattatg atttttoact taacatatat tattaatttt tactgataa  
20461 acccatgtgt taatgtgtag gttgacaaa atgtgtgctg tatgcagtt cgactaggct  
20521 gtttctctct tgaacottta ccaaaatggg ttcaaatgca tcaactatc gagtatgt  
20581 ttctcattaa ttgtattaag aaaatacact acaagttact aacaactatc tcttatcgat

5

10

15

20

25

30

35

40

45

20641 aaacatgaac tcgtttcagg taccttcaa tggattogat tcaaaagtac ctttggccac  
 20701 gattgtgtaa acttcccca gaaactttgt aagttcaaat gtttttctc aattttaagaa  
 20761 gccaaactatt tttagcccat ttgaacacat attacctaat ttattttcta aatattttta  
 20821 cagatatgaa ctacattatc aaatgataac atttggaaa gtgtgggta cttttttctt  
 20881 ttttatatta atgaataaaa taatattggt ggtttaatca aattttgtca actttagggt  
 20941 ttctgcacaa aaactattcc taattgtaat gcattgtcca tgaagtccga atgcaaat  
 21001 ttgtcaagtg catatgtcag gtacaattct ttttctctt cctacttga tacttagata  
 21061 taacttaatt tgttaattcc ataaatatta aagaaaaatc ttagaataat cataaaaaat  
 21121 aattgctaaa cgtctcagct attttatata ataaattttc taaatattga gagtgaatt  
 21181 gagttttaat aattacatta tatatataaa tatataatgt tagaattgac aaattgtgtt  
 21241 tttttttaat agttctaaag ttcttctcga gagtccagaa gaaaagtgc atgagcctaa  
 21301 tacttttatg aatgcacatt ctcaagacgt tgcgttagat atgacatcaa aatataattt  
 21361 ggtagaagaa tgtgtttctt ctggatgtag cgtatcaagt atatgttata agccactagt  
 21421 tgagtttctt tcgtcccca gagcgaaat tcccgagtc acagacattg aagattgttc  
 21481 attcatgaat ctttatcagt catatgtcag ttgttctaaa attgtatttt actgtgatgc  
 21541 attgaagaaa agtgtagaag atgcacttgt aataagtggc aggatgagca gttctgatga  
 21601 agaaatatca aaagcattag tgattccac tctgtaaaat gcattgcttc ctatcaaac  
 21661 acctcggaat atgaagtatt ataactgact aagaactgaa catgtgtgtg aagattcttt  
 21721 atgtaataac tgattatacc atataattta tatgcatttt ttgggaatat ataatctaat  
 21781 actgtgtttt ttgtagtta tgtgttctc gataatcatg agctgtcata cgtatgaagt  
 21841 atcacatac ttaagctaac aaaaaaatgc aactcttttg tatataatc tagaaaaagt  
 21901 ttgtgtttt ttacatatat tatatagtt gagagaagaa aacttgatga tccaaagctt  
 21961 tacttctctg cgaattggca accagggtata atacaagcat aatttatcat tgttccata  
 22021 actataaact aaatttttca ttogaataat ttttagtga aacatcatcc cgtgtctgtc  
 22081 caccaaagaa aaagtgtagt tctgtaggat caaagctttg caagataaag aattgttcat  
 22141 attgttggac tatacagaaa caaaactcca acatttttgc cggaaacaat ttggttaaca  
 22201 aaatttaciaa ttgatattt taacattggt gacttgaac tcacataaat tcaattgtatc  
 22261 agattccatg tagaacagca atgcgagggg cctttccact taatggaaca tacttccaaa  
 22321 ccaatgagge aagcattttt tcttataatt ttttgtctga gtttttactt aatgttttta  
 22381 aagagaacac aatggtttat ttttccaggt ttttgtctga catgagacaa gcttaaaccc  
 22441 cattgtcttt cgtagggagt tgtgtaaggg actagaaaaa cgtgcactat attgtgtgtc  
 22501 aaacgtgaca tctattttta aacttttaga cacaagaagg attgaaattt gcttttggac  
 22561 aggttaacaaa cataaatata tattaattt tttgttgaat tatgaagtta aaataactgt  
 22621 ggaattgtgt gtggtgtgtg gcagggtttt tatgttttag agcatttgat cgaagcaaac  
 22681 gagatccaaa agagcttgtc cgcagctcac acactccacc tgatgagaga ggcccaaggt  
 22741 ttatgagta tgatgatata tttgttctt ttattctttt tggctagtt agcaaatatt  
 22801 ttaaacgaac gaattctttc ttataataac aagcattgca acgattgagt aaatgcaggt  
 22861 acgtattggt tcttgattta aatgcagtga cattataatt atttccaaag gcttttctat  
 22921 atagtagtgt tggatgaaaa agaagagagc ccaagagttg gttctattgt tatgctcggg  
 22981 ttgttgattgt gtggcagcgt cattcaaaac cgagtttatg tttctgggtg gaaggtctct  
 23041 gagtgagga ttccaggaac tgbtctaat cttcttccca cttgtgtgtg caacttttat  
 23101 tttctctgtt ttataagcaa gcttatatgt atcaatgata cagtatcatc tattgttcaa  
 23161 aaaaattgga attaatattc tctctgtctc aacatctttg ggtcgatcgt tattcgatga  
 23221 cagtagcaac tagcagctct cttgtgatat atctagcca agcagctcca aaactttttt  
 23281 tacttctgatt gttgtcagta tttctgtctc agacgttttt agcaaaaaag tttctcaggt  
 23341 gataaaaatta ggttaaaaa agtatgactc tgtctttaag actcagtttc agatagtaat  
 23401 acaaaaatta cataaacaaa gagtggtcac agacgtgtat ctgtaagtgt tgtcagagat

SEQ ID NO:17

**RICE(*Oryza sativa*) DMT1**

**>DMTRICE (1DMTRICE) ;**

MQDFGQWLPQSQTTADLYFSSIPISQFDTSIETQTRTSAVVSSEKES  
5 ANSFVPHNGTGLVERISNDAGLTEVVGSSAGPTECIDLNKTPARKPKKKKHRPKV  
LKDDKPSKTPKSATPIPSTEKVEKPSGKRKYVRKKTSPPGQPPAEQAASSHCRSELK  
SVKRSLEDGGEVLQESTQSGSQVPVAEICTGPKRQSIPSTIQRDSQSQLACHVVSST  
SSIHTSASQMVNAHLFPPDNMPNGVLLDLNNSTSQLQNEHAKFVDSPARLFGSRI  
RQTSKGNSLLEIYAGMSDRNVPDLNSSISQTHSMSTDFAQYLLSSSQASVRETQM  
10 ANQMLNGHRMPENPITPSHCIERAALKEHLNHVPHAKAAVMNGQMPHSYRLAQ  
NPILPPNHIEGYQVMENLSELVTTNDYLTASPSQTGAANRQHNIGDSMHIHALD  
PRRESNASSGSWISLGVNFNQNNGWASAGAADAASSHAPYFSEPHKRMRTAYL  
NNYPNGVVGHFSTSDLSNNENENVASAINSNVFTLADAQRLIAREKSRASQRM  
ISFRSSKNDMVNRSEMVHQHGRPHGSACRESIEVPDKQFGLMTEELTQLPSMP  
15 NNPQREKYIPQTGSCQLQSLEHDMVKGHNLGELHKQVTSPPVVIQSNFCVTPP  
DVLGRRTSGEHLRTLIAPTHASTCKDTLKALSCQLESSRDIIRPPVNPIGPSSADV  
RTDNHQVKVSEETVTAKLPEKRVGRPRKELKPEGKPKRGRPRKGKVVGGELA  
SKDSHTNPLQNESTSCSYGPYAGEASVGRAVKANRVGENISGAMVSLDSDIVI  
QKIKVLDINKSEDPVTAEPHGALVPYNGEFGPIVPFEGKVKRKRRAKVLDLPVT  
20 ALMWKLLMGPDMSDCAEGMDKDEKWLNEERKIFQGRVDSFIARMHLVQGDR  
RFSPWKGSVVDVGVFLTQNVSDHLSSAFMALAAKFPVKPEASEKPANVMFH  
TISENGDCSGLFGNSVKLQGEILVQEASNTAASFITTEDKEGSNSVELLGSSFGDG  
VDGAAGVYSNIYENLPARLHATRRPVVQTGNAVEAEDGSLEGVVSSENSTISSQN  
SSDYLFHMSDHMFSSMLLNFTAEDIGSRNMPKATRITTYTELLRMQELKNKSNETI  
25 ESSEYHGVPVSCSNNIQVLNGIQNIGSKHQPLHSSISYHQTGGVHLPIVHASDLE  
QSVYTGLNRVLDENVTQTSYYPSPHPGIACNNETQKADSLSNMLYGIDRSKDTTS  
LSEPTPRIDNCFQPLSSEKMSFAREQSSSENYSRNEAEAAFKVQHGTSNVQGDN  
TVRTEQNGGENSQSGYSQQDDNVGFQTATTSNLYSSNLCQNQKANSEVLHGVSS  
NLIENSKDDKKTSPKVPVDGSKAKRPRVGAGKKKTYDWDMLRKEVLVYSHGNKE  
30 RSQNAKDSIDWETIRQAEVKEISDTIRERGMNMLAERIKDFLNRVLDHGSIDLE  
WLRYVDSKAKDYLLSIRGLGLKSVCEVRLTLTHHMAFPVDTNVGRICVRLGW  
VPLQPLPESLQLHLEMYPMLENIQKYLWPRCLKLDQRTLYELHYQMIFGKVF  
TKSKPNCNACPMRAECKHFASAFASARLALPGPEEKSLVTSGTPIAAETFHQTYIS  
SRPVVSQLEWNSNTCHHGMMNRQPIIEEPASPEPEHETEEMKECAIEDSFVDDPEE



IPTIKLNFEETQNLKSYMQANNIEIEDADMSKALVAITPEVASIPTPKLKNVSRRLR  
TEHQVYELPDSHPLLEGFNQRPDDPCPYLLSIWTPGETAQSTDAPKSVCSQEN  
GELCASNTCFSCNSIREAAQKVRGTLLIPCRRTAMRGSFPLNGTYFQVNEVFADH  
DSSRNPIDVPRSWIWNLRPRRTVYFGTSIPTIFKGLTTEEIQHCFWRGFCVVRGFDRT  
5 SRAPRPLYARLHPASKITRNKKSAGSAPGRDDE

SEQ ID NO:18

>DMTRICE novel 723 amino acid NH2 terminus;

MQDFGQWLQSQQTADLYFSSIPISQQFDTSETQTRTSAVVSSEKESANSFVPHNGTGLVERISNDAGLTVGVSSAGP  
TECIDLKNTPARKPKKKHPRKVLKDDKPSKTFKSATPIPTSEKVEKESGRKYVVRKKTSPQPPAEQAASSHCRSELKS  
10 VKRSLDFGGVQLRSTQSGSQVFVARICTGPKRQSIPTIQRDSQQLACHVVSSTSSHTSASQMVNHLFPDMPNG  
VLDDLNSTSQLQNEHAKFVDSPARLFGSRIQTSGKNSLLEIYAGMSDRVVDLNSISQTHSMSTDFQAYLLSSQAS  
VRETQMANQMLNGHRMPENPITPSHCIERAALKEHLNHVPAKKAAMNGQMPHSYRLAQNPLIPNHIKGYQVMENLSEL  
15 VTTNDYLTASFPSTQGAANRQHNIGDSMHIHALDPRRESNASSGSNWSLGVNPNQKNWASAGAADAASSHAPYFSEPH  
KRMRTAYLNNYPNGVVGHFSTSDLSNNENENVASAINSVFTLADARLIAREKSRASQRMISFRSSKNDMVNRSEMV  
HQHGRPAFHGSACRESIEVFPKQFGLMTLELTQLSPMPNNFQREKYIETQTSQQLSLEHDMVKGHNLAGELHKQVTSPO  
VVIQSNFCVTPDPVLGRRSTGEHLRLTIAPTHASTCKDTLKALSCQLSSRDIIRPPVNPFIQSSADVPRTDNHQVKVSE  
ETV

SEQ ID NO:19

>DMTRICE nucleotide sequence from PAC PO489G09;

10261 aaatattgct taaatggata taaagtgaa aaatgtactt gaggggaagt gtagggtgca  
10321 gtgggggtccc acaatttttt ttcaactagt caacttagt tatataattt ttgocgaaga  
10381 ggacaaaaggc gctccgtgta attttgagta agggcgccgg ggaattttat ttgtgtaaag  
25 10441 gaacctagcca agaaaagcat gatagtgcac atgtatccct tcttttttct ttctttttgtt  
10501 ttcataaactg tcttacagaaa ttctatgttg gctgggtgaca cttgtctcac tcattatttt  
10561 gtatatatttg actaaatgca acgtgtgtgt gctcgtgagt ttatatattg ttttaacgat  
10621 tcttcattga ctgtatgtat ttgatgttga tacctctggg tgtcttattt tatagggtga  
10681 tgcctggagg ccacatagga ggccctgtgtg atccaaagtgt gctgctcctg agttgaaatt  
30 10741 gcatagccat atagcaacta ctgggtgtaaa cttgagagat gaagtagtga aaggaaatat  
10801 gcaggatttt ggacaatggc tgcctcaact tcagaccact gccgatctat attctccag  
10861 tattccaata ccatcacagt tcgatacttc catagagacg cagactagaa cttctgcagt  
10921 tgtatcgtca gagaagaagt ctgctaattc gttcgtccct cataatggtc ctgggctgtg  
10981 tgaacgcatt agcaatgatg ctgggctaac tgaagttagt ggaagtagtg ctggaccacac  
35 11041 tgaatgtatt gaactgaaca agacaccagc acggaaaacc aagaagaaaa agcacaggcc  
11101 aaaggtgcta aaggagcata aaccatcgaa gacacctaaa tctgtactcc caataccttc  
11161 aacagaaaaa gtgaaaaaac catctggaaa gagaaaatat gtcgcgacaa agacatctcc  
11221 agggccaaact cctgcagaaac aggcagctag ctccactactc agatctcgagc tgaagttagt  
11281 taaacgaagt ttggactttg gtggagaagt actgcagagc agtacaacat cttgatctca  
40 11341 agttcgggtg gcagaaatat ctactgtgct caagcgtcaa tcaatacctt ctacatctca  
11401 aagagattcg caaagccagt tggcttgcca cgtgggttct agccaccagt caattccacac  
11461 ttcagctagt cagatggtta atgcacattt gtttctcctc gataacatgc caaattggat  
11521 atttggtgac ctcaataaatt ctactagtca gttacaaaac gaacatgata aatttggtgta  
11581 cagtcgggca cgtctttttg gttccagaaat aagacagaca tcaggtaaaa attcttttgt  
45 11641 agaaattctat gctggcagtg cagatagaaa tgtacctgat ctcaacatct caatcagta  
11701 gaagcagatg atgtctactg attttgtcca atacttgctt tcatctcact aagctctcgt  
11761 aagggaacaa caaattggcca atcagatgct taatggctat aggatgccag aaaatccaat

5

10

15

20

25

30

35

40

45

50

11821 tacacctagt cattgtattg aaagggctgc attgaaggaa cacttgaatc atgttcotca  
11881 cgcacaaagcc gcagtgatga atggcccaat gccccatagt tacaggttgg cgcacaaatcc  
11941 catctcatcct ccaaatcata tgaagggtta tcaagtgatg gaaaatttga gtgaacttgt  
12001 cagcagcaaat gactatctaa ctgctagtc tttcagtc aaactgagctg caaataggca  
12061 gcataatatt ggtgactcca tgcataata tgcattggat cctagaagag agagttaatgc  
12121 tccaagtggt tcttggtgat cattaggtgt gaactttaac caacaaaata atggatgggg  
12181 atctgcaggt gctgcgatg ctgcgatgc acatgcccc tatttttcag aacctcaca  
12241 aagaATTGagg acagcttatc ttaacaata tccaaatgga gtctggggac atttttctac  
12301 ctcatctacg gattttgcaa ataagtgaag tgaataatgt gccacagcaa tcaactcaaa  
12361 cgttttttacc ctgtctgatg cacaagatt gatagccgtg gagaatcac gagcttccca  
12421 aagaatgac agtttttagat catctaaaaa tgatatgggt aacagatcag aaatgggtcca  
12481 tcaacatggc agacctgctc cgcattggctc tgcattgcagg gactctattg aagtaacctga  
12541 caaacagttc gggctcatga cagaagaact cacacaatta cctagtatgc caaataaccc  
12601 acaaaaggaa aaatatatc cgcacaaactg aagttgccaa ctccagtcct tggaaactga  
12661 catgggttaa gggcataact tggcagggtga attgcataag caactaaact caccctcaagt  
12721 tgttatttcag agcaattttc gtgttaccct tctctgatgt ctcggcagaa gaaccagtgg  
12781 ggagcattta agaaccctta tagctccaac acatgcctgt acatgtaagg acactctgaa  
12841 agctttaagt tgtcaactgg agagttctag agacattatt aggcctcctg tcaatccat  
12901 agggcccatcc ctgctgatg ttccaagaac tgataacct caagtcaagg tttctgaaga  
12961 aacccgttaca gccaaactcc ctgagaagcg aaaagttaga cgtcccgaaa aagagttcaa  
13021 accctgtgag aaacccaaac ctagaggcgc tccaaggaa ggaacaggtg ttgggggaga  
13081 acttgcatca aaggatagtc acactaatcc attgcaaaat gagagtact catgttctta  
13141 tggctcctat gcagggggagg ctctgtgtgg aagagcagtt aaagcaataa gagttggaga  
13201 aaacatttct ggagctatgg tatccctact ggattcttta gatattgtta ttcaaaagat  
13261 aaaggtcttg gacataaaca aatcagaaga cctgttgaca gctgaacccc atgggtgctc  
13321 tgtcccttac aatggagaat ttggtcctat tgttctcttt gagggaaga tgaagaagaa  
13381 agctctcga gccaaagtgg atcttgacc tgaactctt ttaattgtga agttactaat  
13441 gggaccagat atgagtgatt gtgtgaagg tatggataag gataaagaga aatggctaaa  
13501 tgaagaaga aaaatattcc aagggtgtgt tgattcattt attgtcgaa tgcattcagt  
13561 tcaaggtatt tctatcattt taaaattgt ttccctaact gaacatgatg gcttccatct  
13621 tgtgattgct gccctcacat tagtgaatgg tctcaaatct tcaatattta ctgtgtacc  
13681 aaactcctat tcttcatccc aatatattca tgtttgtact cgtactgtcc cattagactt  
13741 gactgtgct gtgaagatca acaactttac ttttaggtt acctctatgt ttgcaggaga  
13801 tcggcgcttt tctccttgga aaggatcagt tgtagattct ctgagtgagg tatttcttac  
13861 acagaaatgt tcggaccatc tttccagggt aataatgct agagctactt tgaanaatgt  
13921 gacttgactt gcattgtgag gttatgtgt ttttctgtct gactattccc tttttttca  
13981 gctctgactt tatggtcttt gctgcacaa tctctgtaa gccacagcc ctgaaaaaac  
14041 ccgcaaatgt gatgtttcat acaatttcag aaaaatgtga ttgttctggg ttgtttggt  
14101 attctgtcaa gctacagggt gagatcctg ttccaggagg cagcaacaca gcagcctctt  
14161 tcatcacaca ctaggataag gaaggaaagt acagtgtgga attgcttgga agttcttttg  
14221 gggatggagt ggtaggtgca gcaggagttt attctaatat ttatgagaat ctgccagcta  
14281 gactgcatgc tactaggcgt ccagctgttc aaactggaaa cgtctgcgaa gcggaagatg  
14341 ggtcactgga ggtgtgtgt tcatcagaaa actccactat ttcactcaca aattcactag  
14401 attctctatt tccatgtct gatcatatgt ttctagcat gttactaaat ttcaatgcg  
14461 aagacattgg cagcagaaat atgccccaa caacaagaac cacatatata gaacttctac  
14521 gaatgcagga gctgaagaac aagtctaatg aaacattgga atcatcagag tatcatgggg  
14581 tctcagttct atgtagttaac aacattcaag tgcctcaagt aatcacaaat atcgccagta  
14641 aacatcagcc ttacattccc tctatttcat atcacaggac tggccaaagt cactctccag  
14701 acatagtaca tgcagtgat ttggagcaat cagtatacac tggccttaat agagtgtgt  
14761 attctaatgt tacacaaacc agttattatc ctctccctca tctggaaat gctgttaaca  
14821 atgaacaca aaaggctgac tctttaagca acatgttata tggatatagat agatcagata

2

82

5  
10  
15  
20  
25  
30  
35  
40  
45  
50

83

21001 ttgaactttt cattgtcttc tgaatgtatc ttactcttgg atcattaata ttctattttg  
 21061 tcacgatata gtggatatagg acaataaaat catgggaagt atttattttc atcaccaatc  
 21121 tactcatata attttcaaat gacaattata aatatcttaa aaatatattg ttagtgtgtc  
 21181 tgtataaaat aattgtcaca cctcagtcga cagcgacaag aatttgtgtc tacaggctag  
 21241 agtgagtact ctagaagtat cttcatagga atcggaataa aatgccaatg tgaatgaaca  
 21301 aggatatacaa gtataccctc aaaatctcta gagaggattg cgtaaaatg taggtgtaac  
 21361 taaacaattg ttctcatatg agggctttct taaaggaggt acaagactta tcaatatggg  
 21421 taaagtattt ttatccata ggcattgttg gcagaaagct gcttagggta gaatgtact  
 21481 ccctccgtcc cacaataaa gagattttga gtttttgctt gcaacgtttg accactcgcc  
 21541 ttattcaaaa atttttgaaa ttattattta ttttatttgt gactactttt attattcaca  
 21601 gtacttttaag tacaactttt cgttttttat atttgcaaaa aaaattgtat aagaacgagt  
 21661 gtcaaacgtt gtacgcaaaa actcaaaatc ccttatattg tgggacggag ggagtactta  
 21721 tggatgcctt ttttgtccaa gatgtcagta acattttctt tcagggtatg ggatttttac  
 21781 ttcttttttc cctaactttt tcaggatttg tgtgogtgag aggcctttgat aggcacatcaa  
 21841 gagcaaccag accactgtat gcaagactcc actttccagc aagcaaaaat accaggaata  
 21901 aaaaatctgc aggttctgct ccagggaagag atgatgaata ggccactcgg aaaccagaaa  
 21961 aggaataaaa gaggaggtac atatgatctg ccagaagatc actgacctga aatggatcgc  
 22021 tgaccaataa gttgcgttag gcaattcaat tatttctggc catatacatc tgcggaagt  
 22081 tatgaactcc agccactgac gaattcgtgg tgctggtatt cttaggcaac atgatccatc  
 22141 atacagattc tatgcttggt tgttgcaagc aattcttatg cggtagcagt tgcgtcgtgat  
 22201 agggagaaaa ggcattgtcc gggctcagc ggctctaact gtaacttcat atgagtgaaa  
 22261 ccgattgttg tacatgtgaa aagtttgcca tcaaaaatgg tcaattcatg tgttaggtca  
 22321 ttcatgtagt cgaatgtcaa ttaatcatca attattgat ttgattcatt cacaagttaa

SEQ ID NO:20

CORN(ZEA MAYS)DMT.1

>Corn DMT.1 660990 (688512 selclone ID);

EPDDPCPYLLSIWTPGETAQSIDAPKTFCDSGBTGRLCGSSTCFSCNNIREMQAQKVRGT  
 LLIPCRTAMRGSFPLNGTYFQVNEVFADHCSSQNPIDVPRSWIIDLPRRTVYFGTSVPTI  
 FRGLTTEEIQRFCWRFVFCVRGFDRTVRAPRPLYARLHFVSKVVRGKKPGARAEE

SEQ ID NO:21

>Corn DMT.1 cDNA 660990 (688512 selclone ID);

gaaccagatgatccttgtccatatcttctttccatatggaccccagggtgaaactgcacaa  
 tgcgatcgatgcccccaagac  
 attctgtgattcaggggagacgggtagacatattgtggaagtccaacatgcttttagttgcaa  
 caatatacgagaaatgcagg  
 ctcagaaaagtcagaggaaacttttgataccatgccgaacagcaatgagaggaaggtccc  
 cacttaatgggacgtatttt  
 caagttaatgaggtatttctgacaccattgctcaagtcaaaaatccaattgatgtcccacga  
 agttggatttgggacctccc

aagacgaactgtttactttggaacctcagttcctacaatattcagagggtttaacgactga  
 agagataacaacgatgctttt  
 ggagaggatttgtttgcgtgaggggctttgataggacagtgcgggcaccaaggcccttt  
 atgcaagggttgcattttctt  
 5 gtcagcaagggttgtagaggcaaaaagcctggagcagcaagagcagaagaataatagaac  
 attgaagaaatataggggtg  
 ctaaccagatgaggatggatagcccgaatgagatgctgacccaataggctcgcaaatca  
 cctccaaattctaaccaat  
 gacttccatctgtaatgaatggcaataccttgaaaacctgtgatggagatgttttggg  
 10 gacatgatctcttaaattag  
 attcogtctttggtaacagcctagctgttcttgttgagtcgcatattctttattctgaag  
 atcaatatagcaaattggg

SEQ ID NO:22

15 CORN(ZEA MAYS)DMT.2

>Corn DMT.2 371537 (441428 selclone ID);

MITFGKVFTKRPNCNACPMRSECKHFASAFASARLALPAPQEESLVKLSNPFQNSS  
 MHAMNSTHLPRLEGSIHREFLPKNSEPIIEEPASPREERPPXTMENDIEDFYEDGEIPT  
 IKLNMEAFQNLENCIKESNNELQSDDIKALVAIXTEXASIPXPK

20 SEQ ID NO:23

>Corn DMT.2 cDNA 371537 (441428 selclone ID)

tatcagatgattacatttgaaaggtctttgtacccaaaagacagccaaattgcaatgca  
 tgcccaatgaggagtgtg  
 25 caagcattttgcaagtgcatttgcaagtgcaaggcttgcaacttctgtctcccaggagga  
 aagcttagtgaagttgagca  
 atccatttgcctttccagaatagcagcatgcatgctatgaattcgactcacctacctcgcc  
 ttgaggggagtatccattca  
 agggagtcttctcctaagaactcagagccaataatcgaggagcctgcaagtcgaagagag  
 30 gaaagacctccakaaaccat  
 ggaaaatgatattgaagattttatgaagatgggtgaaatcccaacaataaagcttaacat  
 ggaagcttttgacacaaaact

tggagaattgcattaaagaaagcaataacgaactccagctctgatgatattgcaaaagcat  
 tgggttgcattaractgaa  
 rcagcttcsattcctgkaccgaaact

5 SEQ ID NO:24

**Corn(*Zea mays*)DMT.3**

>Corn DMT.3 218853;

MPRKPKRKAPASPARHDPSPPEYP SHASPCSAQCLVVRDALLAFHGFPEEFAAFRVLRLG  
 GLSPNRDPRPSSPTVLDGLVTTLSSQNTTDAISRRAFASLKAAFP SWDQVVD EEGKRLED  
 10 AIRCGGLAATKAARIRSMRLDVRERRGKICLEYLRELSVDEVKKELSRFKGIGPKTVACV  
 LMFY LQKDDFPVDTHVL RITKAMGWVPATASREKAY IHLNNKI PDDLKFDLNCFLVTHGK  
 LCQSCTKKVGS DKRKSNSACPLAGYCCIG EKLQQL

SEQ ID NO:25

15 WHEAT DMT.1

>Wheat DMT.1 614028 (887053 selclone ID);

MRAECKHFASAFASARLALPGPEEKSLVTSGNPIASGSCQQPYISSMRLNQLDWNANAH  
 HILDNRQPIIEEPASPEPEPETAE MRESAIEDIFLDDPEE IPTIKLNFEEPAQN LKNYMQ  
 VNNIEMEDADMSSALVAITPEAAS IPTPRLKNVSR LRTHEQVYELPDSHPLLEGYDQREP  
 20 DDP

SEQ ID NO:26

>Wheat DMT.1 614028 (887053 selclone ID);

tgcccaatgagagctgaatgcaagcactttgcaagtgcatttgcaagtgcactgtctctcctggacctg  
 25 aagagaagagtgttggtacgtcaggaaaccaattgcttcaggagctgccagcagccatacataagtctatgcgtttaaatcaa  
 ctgactggaatgcaaatgccatgaccataattctggacaatgccagccaatcattgaggagccgaagtcggaaaccagaa  
 ccagagactgcagagatgagagagagtgccatagaggatattttctgatgatcctgaagaaatcctacaatcaagcttaatttc  
 gaggagttgacagaatctcaagaattatgatcaagtcaataacattgaaatggaagatgctgatatgtcaagtcgcttggtgcc  
 ataactccggaagctcatctatccgactcctaggtcctcaagaatgtagtcgctaagaacagagcatcaagtcctatgaactgcc  
 30 ggactcacatccactcttgaagatacgaacaaagagagcctgatgaccttg

SEQ ID NO:27

Wheat DMT.2

>Wheat DMT.2 568842 (908118 selclone ID);

NRVDESTVGGADKAASPKKTRTTRKKNTENFDWDKFRQACADGHMKERKSERRDSVDWE  
AVRCADVQRISQAIRERGMNNVLSERIQEFLNRLVRDHGSI DLEWLRI PDPSAKDYLLS  
IRGLGLKSVECVRLTLHHLAF PVD

# 5 SEQ ID NO:28

>Wheat DMT.2 568842 (selclone ID 908118);  
caaacagggtggatgaatctactgtcggaggagcagataaagcagcaagctccaaagaaaaaagaaccacaaagaaaaaa  
aatactgaaaacttcgactgggacaaatttcgaagacaggcctgtgctgatggccacatgaaagaaaggaagtctgaaag  
aagagactctgttgattgggaagcagctacgatgtgcagatgtacaaagaatttctcaggccatccgggaacgaggaatga  
10 ataatgttttatcagaacgaatccaggaattcctgaatcgcttggttagagatcatggaagcattgatcttgaatggtta  
agagatatccccctgactcagcgaaggactacttgccttagcatacgtggactggggctcaaaagtgtgaatgtgttcg  
tctactgacattacatcatctcgctttccctgtwgacac

# SEQ ID NO:29

## 15 WHEAT DMT.3

>Wheat DMT.3 611792 (838515 selclone ID);  
NRKQVNEVFADHKSSYDPIYVAREQLWKLERRMVYFGTSVPSIFKGLTTEETIQCFWKGF  
VCVRGFERETGAPRPLCQHLHVAASKVPRSRNAAAAGLNSDSAKASAP

# 20 SEQ ID NO:30

>Wheat DMT.3 611792 (838515 selclone ID);  
aatcgaaaacaagtaataatgagglatttgcagaccacaaatctagctacgtacccatatacgttgcaggga  
gcagttatggaagttgaaagacgaatggctactttggaacttcagtgccctccatatcaagggtctaacaactgaagaaatata  
gcagtgctcttggaaagatttgcctgtgtgcggggattcgagagggaaccggggcaccaggcctctatgccaacatctgca  
25 cgtgcgggctagcaaaagtcggagatcacgaacgcggcagcagctgggctgaactcggattcagcaaaagcattcggctcca  
tgagtatcatcacaccggctatgcactgtgcatgggtacgtagtgtgttctgccggcgcwacagccgtyttgttaggaata  
aaccscgtgcgaaraagaattatcatccagttgtytgagtgtatacttctgctgtagkacctttttTaaatccctgtgagctyattg  
tactctgaatttacttccgaccagttatccgcttgcgaaraggcctttgttatgkaccggcactctgtgtatatacatcatgttcttc  
traaaaacttgtcttccakacgaccttacgt

30

# SEQ ID NO:31

## Wheat DMT.4

>Wheat DMT.4 615131 (861906 selclone ID);

09840743-042301



MRSECRHFASAFASARIALPAPQEKSLVMSSNQFSFQSGMPTPYSTVLPQLEGSAGGQD  
FCTNNSEPIIEEPASPAEECPETLENDIEDYDPTGEIPLIKLNLQAFQNLNICKES  
NMDLGSDDIAKALVAVSTGSASIPV

5 SEQ ID NO:32

>Wheat DMT.4 615131 (861906 selclone ID);

tacttttggaagggtgttctgtacaaaaaacaagccaaattgcaatgcttgtccaatgag  
aagcgaatgcaggcatcttcgcaagtgccttcgcaagtgcacggcttgcacttctcgacc  
tcaggagaaaaagtttggtgatgtcgagcaatcaattcagttccagagtgggtggcatgcc  
aactccatactcaactgtgcttctcagcttgagggaagtgccaggacaggatttttg  
cactaacaattcagagccaattattgaggagccagcaagtccagcacgggaagaatgtcc  
agaaactcttgaaaaatgata  
ttgaagattacgatccagatactggtgaaatcccactaattaagcttaacttgcaagctt  
ttgctcagaacttgaaaaactgcattaaagaaagcaatatggatcttgggtctgatgata  
tcgcgaaagcacttgttgctgttagcactggatcagcttcaattcctgtccc

SEQ ID NO:33

**Soybean(*Glycine max*)DMT.1**

>Soy DMT.1 449122 (557119 selclone ID);

MDSLDWDVAIRCADVSEIAETIKERGMNRLADRIKNFLNRLVEEHGSIDLEWRDVPDPK  
AKEYLLSIRGLGLKSVECVRLTLHHLAFVDTNVGRIAVRLGWVPLQLPLPESLQLHLL  
LYPVLESIQKYLWPRCLKLDQETLYELHYQMITFGKXFCTKSKPNCNACPMRSECRHFAS  
AFASARFALPGPEQKSIIVSTTGNVINQNPSEIISQLHLPPENTAQAEDIEQLTEVSRQL  
ESKFEINICQPIIEBPRTPEPECLQESQTDIEDAFYEDSSEIPTINLNIIEFTLNLQN

SEQ ID NO:34

>Soy DMT.1 449122 (557119 selclone ID);

aataaaatttaakagcaaggacaagaaaaagagaaaaagatgattgactgggatgttaagaattg  
aagcacaggctaaagctgggaaaagagaaaagacagataaacaccatggattcttggactgggatgctgtgagatgtgcagat  
gtcagtgaaatcgctgagaccatcaaaaggggcatgaacaacaggcttcagatcgtattaagaatttctaatacgtattgtt  
tgaagaacatggaagcattgacctgaatggcttagagacgttcacactgacaaagcaaaagataactgctcagcataagagga  
ttgggactaaaaagtgtggaatgtgtgcggctttaacactgcaccacttgccttccggtagacacaaaatgtcggacgatagca  
gtacgactgggatgggtccctctacagccactgcctgagtcactgcagttgcatctctagaattgtaccagtggttgagtaata

caaaaatatcttggcctgactatgcaagctagatcaggaaacactatatgagctacattaccagatgattacatttggaaaggkc  
ttctgtacaaaaagcaaaccaattgtaatgcatgcccaatgagaggagaat

SEQ ID NO:35

5 **SOYBEAN(*GLYCINE MAX*)DMT.2**

**>Soy DMT.2 387990 (473695 selclone ID);**

MRMTIDLVSQQSLTARLQLSILKDKLKIQCRKARGLDVGRNESSKIDSSPVKLRSREHGK  
EKKNFNDWDSLRIQAEAKAGKREKTENTMDSLWDVAVRRADVSEIANAIKERGMNNMLAE  
RIQSFLNLLVDKHGGIDLEWLRDVPDQAKEFLLSIRGLGLKSVECVRLTLHLHLAFVVD  
10 TNVGRIAVRLGWVPLQPLPESLQLHLELYPVLESIQKYLWPRICKLDQRTLYELHYQLI  
TFGKVFCTKSK

SEQ ID NO:36

**> Soy DMT.2 387990 (473695 selclone ID);**

15 gaaaagatagggatcattctcagatagcaactcagaaatagaagacctgtctagcgctggc  
aagtacaatagttattataa  
tagaatttctttcagtgagcttttagaaatggcaagtccaacatggtgcatgaagttaa  
cagtcaaagaagcaaatcaa  
ctgagaacttaggagatacatgtgatcagtcctatagacatgaagcatgacaacctggcag  
20 aaaacttggaataatcggat

gttactcaaggctccgcagaagcaccatcaccaatggatatactttttaaataaaccca  
aactcaggagtacttgaggt

taactgttatgatcctctcaaaatagaagtcacatcaagtggctcctcaaagggtaaaga  
tgagaatgacaatagatcta

25 gtttcccaacagagctctgactgccaggctgcaattgtccattctcaaggacaaactgaag  
atccaatgacGaaaagcaag  
gggactagatttttgtaggaatgaaagcagtaagatagattcttccctgtaaaattaag  
gagcagggagcatggaaaag

agaaaaagaataaactttgattgggatagtttaagaatacaagcagaagctaaggcagggg  
30 aaagagaaaaagacagagaaac  
accatggactccttgactgggatgctgttagacgcgcagatgtcagtgaaattgccaat  
gcaatcaaagaaaggggcat

gaacaacatgcttgctgaacgtattcagagtttctgaatctatttggttgacaagcatgg  
 gggcatcgatcttgagtggc  
 tgagagatgttccacctgatcaagcaaaagaattcttgctcagcataaggggattgggat  
 tgaaaagtgtggagtgtgta  
 5 cgactcttaacactacaccatcttgcccttccgggtggacacaaatgttggaagttagca  
 gtaagattgggattgggtgcc  
 tctccagccactgccagagtcactacagttgcatcttctagaattgtaccagtggtgga  
 gtccatacaaaaatatctct  
 ggccccggctctgcaagctagaccaaagaacattgtatgagctgcattaccagctgatta  
 10 catttggaaggtcttctgt  
 actaaaagcaagcc

SEQ ID NO:37

SOYBEAN(*GLYCINE MAX*)DMT.3

15 >Soy DMT.3 657152 (546665 selclone ID);  
 INQAEQQTEVIRQLEAKSEINISQPIIEEPATPEPECSQVSENDIEDTFNEESCEIPTI  
 KLDIEEFTLNLQNYMQENMELQEGEMSKALVALHPGAACIPTPKLKNVSRRLRTEHYVYEL  
 PDSHPLNLGWNKREPDDPGKYLLAIWTPGETABSIQPPESKCSSQECCGLCNENECFSC  
 NSFREAXFXDSXRDTPDTMSNSXXXGAFH

SEQ ID NO:38

>Soy DMT.3 657152 (546665 selclone ID);

tataaaccaagcagaacttcaacaaacagaagtgatcaggcaactagaagcaaaatctga  
 aatcaacatcagccaaccta  
 25 ttattgaagagccagcaactccagagccagaatgctcccaagtatccgaaaatgatatag  
 aggataccttcaatgaggaa  
 tcatgtgaaattcccaccatcaaactagacatagaagagttcactttgaaactacaaaac  
 tatatgcaagaaaacatgga  
 acttcaagaaggtgaaatgtcaaaggccttggtgctctacatccagggtgctgcatgcat  
 30 tcctacaccaagctgaaga  
 atgtgagccggttgccaacagagcattatgtttatgaactccctgattccacatccccttc  
 tgaatgggtggaacaagcga  
 gaacctgatgatccaggcaaatccttctagctatatggactccaggggagacagcagat  
 tctatacagccaccagaaag

caaatgcagctctcaggaatgtggccggctctgtaatgagaatgaatgtttttcatgcaa  
cagtttccgtgaagcaagggt  
tcacagatagttcgagggacactcctgataccatgtcgaacagctwtgaragggag

5 SEQ ID NO:39

SOYBEAN(*GLYCINE MAX*)DMT.4

>Soy DMT.4 432980 (663678 selclone ID);

EAASIPMPKLNVSRLRTEHCYELPDTHPLLQGWDTREPDDPGKYLLAIWTPGETANSI  
QPPESKCSSQEBCGQLCNENECFSCNSFREANSQIVRGTLV

10

SEQ ID NO:40

>Soybean DMT.4 432980 (663678 selclone ID);

agaagctgcttccattcctatgccaagctaaagaatgtgagccgattacgaacagagca  
ttgtgtttatgaactccag

15

atagcatcctcttccaagggtgggacacacgagagcctgatgatccaggcaaatatc  
ttcttgctatatggactcca  
ggtgagacagcaaatctatacagccaccagaaagcaaatgcagctctcaagaagaatgt  
ggccaactctgtaatgagaa  
tgaatgtttctcgtgcaacagtttccgtgaagcaaatctcagatagtttagagggacact

20

cctgggtcgaatgcttatca  
aaatcattgttttaaccatattgtagcttactaattcttatacattatgggaacaggggag  
ggaatacatctccatagaaa  
ttcaagcattataatagactgacttgaatttatgataaatatgagcagataccatgt

SEQ ID NO:41

>Medicago 6654943;

MELQEGEMSKALVALNQEASYIPTTKLKNVSRRLRTEHSVYELPDSHPLLEGWEKREPDDP  
GKYLLAIWTPGETANSIQPPDRRCSAQDCGQLCNEEECFSCNSFREANSQIVRGTILIPC  
5 RTAMRGSFPLNGTYFQVNEVFADHESSLNPISVPRSLIWNLDRTTVHFGTSVTSIFKGLA  
TPBIIQQCFWRGFVCRVSFERSTRAPRPLMARLHFPAS

SEQ ID NO:42

>Medicago 6654943 EST306265

10 GAGAACATGGAACTTCAAGAAGGTGAAATGTCAAAGGCCTTGTTGCTCTAAATCAAGAA  
GCTTCTTACATTCTACAAAGCTGAAGAAGCTGAGTCGGTTGCGCACAGAGCATCTT  
GTTTATGAATCCCAGATTCTCATCTCTTCTGGAAGGGTGGGAAAAGCGAGAACCTGAT  
GATCCAGGAAAAATACCTTCTAGCTATATGGACGCCAGGTGAGACTGCAAATCTATACAG  
CCACCAGACAGAAGATGCAGCGCTCAAGATTGTGGCCAACTCTGTAATGAGGAGGAATGT  
15 TTTTCGTGCAACAGCTTCCGTGAAGCAAATTCACAGATAGTTCGAGGGACAATCTCGATA  
CCATGTCGAACAGCTATGAGAGGGAGCTTCCGCTAAACGGAACCTATTTTCAAGTCAAT  
GAGGTTTTTGCAGACCATGAATCAAGTCTTAATCCGATTAGCGTTCCCGAGAAGTTTGATA  
TGGAACCTTGATAGGAGGACAGTGCATTTTGGAACCTCCGTAACAAGCATATTCAAAGGT  
TTAGCAACACCAGAAATTCAACAGTGCTTCTGGAGAGGGTTTGCTGTGTGCGGAGCTTT  
20 GAAAGGTCAACGAGAGCACCCCGTCTTTAATGGCCAGACTGCATTTCCAGCAAGC

SEQ ID NO:43

>Tomato 12624037;

MELQEGEMSKALVALNQEASYIPTTKLKNVSRRLRTEHSVYELPDSHPLLEGWEKREPDDP  
25 GKYLLAIWTPGETANSIQPPDRRCSAQDCGQLCNEEECFSCNSFREANSQIVRGTILIPC  
RTAMRGSFPLNGTYFQVNEVFADHESSLNPISVPRSLIWNLDRTTVHFGTSVTSIFKGLA  
TPBIIQQCFWRGFVCRVSFERSTRAPRPLMA

SEQ ID NO:44

30 >Tomato 12624037 EST469495

GCTTGAGAAAGGAAGTCCAATCAAAGAGTGGGAAAAAGAAAGCAAGGATGCAATGG  
ACTCATTGAACTACGAAGCAGTCAGAAGTGCAGCAGTTAAAGAAATTTCTGATGCTATTA  
AGGAACGAGGGATGAACAACATGCTGGCAGAGCGAATTAAGGACTTCTCTGATAGACTGG

TGAGGGATCATGGAAGTATTGACCTAGAATGGTTGAGAGATGTGGCCCCAGACAAAGCGA  
 AAGAGTATCTTTTGTAGTATTCGTGGACTGGGTCTGAAAAGTGTAGAATGTGTGCGGCTAT  
 TAACACTTTCATAACCTTGCTTTTCCAGTTGACACAAATGTTGGACGAATAGCTGTGAGAT  
 TAGGATGGGTTCCCTCTCCAACCACTTCTGTAGTCCCTGCAGTTGCATCTTCTTGAAGTGT  
 5 ATCCAATTCTGGAGTCAATTGAGAAGTATCTCTGGCCACGACTCTGCAAGTTCGATCAGA  
 GAACACTGTATGAGTTGCACTACCACATGATTACCTTTGGAAAGGTTTCTGCAACCAAAA  
 GTAAGCCTAACTGTAATGCATGCCCACTGAGAGCTGAATGCAGACACTTTGCTAGTGCTT  
 ACGCAAGTGCAAGACTTGCCCTTCTGGCCACAGAGAGAAGAGTATAGTGAGTTTACGACG  
 TTCCGATCCCTAGTGAGGGAATGCAGCTGCCGCATTCAAGCCCATGCTATTACCCCCAG  
 10 AGCTGAAGTAGGGATGGCGTACCCATATGCTCCAATTG

SEQ ID NO:45

>Barley 13256964;

MASETETFAFQABINQLLSLIINTFYSNKEIFLRELISNASDALDKIRFESLTDKSKLDA  
 15 QPELFIHIIPDKATNTLTLLIDSGIGMTKSDLVNNLGTIARSGTKDFMEALAGADVSMIG  
 QFGVGFYSAYPCARVXVTSKHNDDEQYGEXQAGWLLYCGHVILLESFPFGVLRSPSTS  
 RTNSWSTLERRAFKDLGKNT PSS

SEQ ID NO:46

20 >Barley 13256964 HVSMEI0014B12F  
 CGAGAACCCCGCTCCAAAGCCCTAACCTTAGGCCATCCCCTCTCCCTCCCCTCAACCCCTC  
 GTCGACTCCGCGCGCGCTGCGTTCCAGGAGCTTCCGCTGCCGGCGGCCCATGGCCTCA  
 GAGACCGAGACCTTCGCTTCCAGGCGGAGATCAACAGCTGCTCTCGCTCATCATCAAC  
 ACCTTCTACTCCAACAAGGAGATCTTCTCCGCGAGCTCATCTCCAACGCCTCCGATGCG  
 25 TTGGATAAGATCAGGTTTGAGAGCCTCACTGACAAGAGCAAGCTGGATGCTCAGCCAGAG  
 CTGTTTATCCACATTATCCCTGACAAGGCCACCAACACACTCACCCCTTATCGACAGTGGC  
 ATTGGTATGACCAAGTCAGACCTCGTGAAACAACCTTGGTACCATTGCAAGGTCTGGCAC  
 AAGGATTTTCTATGGAGGCATTGGCTGCTGGTGCCGATGTGTCCATGATTGGTCTGGT  
 GTTGGTTTCTACTCTGCTTACCTTGTGCTGAGAGAGTCGNTGTGACCAGCAAGCACAAC  
 30 GATGACGAGCAGTATGGGGGGGAGTNCAGGCTGGGTGGCTCTTTACTGTGGACACGTG  
 ATACTCTTGAGAGCCCCCTTTGGAGGGGTACTAAGATCCCCCTTACTCTCAAGGACGAAC  
 AGTTGGAGTACCTTGGAGAGGCGCGCTTTAAGGATTTGGAGAAAAACACTCCGAGTCA  
 TAACCTTTTCATCTCCTCTGGACGGGGAACCCCTGAAAAGGAATTTTGGCGCTGGAAA

GTGGGTGAAAAATGGGTTCTTGGGGGGCCCGTTGAGGGATTGTTGGTCACATAAACAA  
ACTATCGTCTTCTATCTTAGCACCTAATAGTCCTTCACATGAG

SEQ ID NO:47

5 >Corn 1BE511860;

LLEGFEQREPDPCPYLLSIWTPGETAQSIDAPKTFCDSETGRLCGSSTCFSCNNIREM  
QAQKVRGTLLIPCR TAMRGSFPLNGTYFQVNEVFADHCSSQNPIDVPRSWIWDLPRTTVY  
FGTSVPTIFRGLTTEBIIQRCFWRGFVCVRGFDRTVRAPRALYAR

10 SEQ ID NO:48

>Corn 1BE511860 946063H01.Y1 946 -

TATGAAGTCCAGATTACACGCCTCTTCTGGAAGGATTGGAACAGAGAGAACCAGATGA  
TCCCTGTCCATATCTTCTTTCCATATGGACCCAGGTGAACTGCACAATCGATCGATGC  
CCCCAAGACATTTCTGTGATTGAGGGGAGACGGGTAGACTATGTGGAAGTTCAACATGCTT  
15 TAGTTGCAACAATATACGAGAAATGCAGGCTCAGAAAGTCAGAGGAACACTTTTGATACC  
ATGCCGAACAGCAATGAGAGGAAGCTTCCCACTTAATGGGACGTATTTTCAAGTTAATGA  
GGTATTTGCTGACCATTGCTCAAGTCAAAATCCAATTGATGTCCCACGAAGTTGGATTG  
GGACCTCCCAAGACGAAGCTTTTACTTTGGAACCTCAGTTCTTACAATATTGAGAGTTT  
AACGACTGAAGAGATACAACGATGCTTTTGGAGAGGATTTGTTTGCCTGAGGGGCTTTGA  
20 TAGGACAGTGCGGGCACCAGGGCCCTTTATGCAAGG

SEQ ID NO:49

>Cotton 11206330;

MQGNMELQEGDLSKALVALNPDAASIPTPKLKNVSRLRTEHYVYELPKHPLLKQMEKRE  
125 PDDPSPYLLAIWTPGETANSIQPPEQSCGSQEPGRLCNEKTCFACNSVREANTETVRGTI  
LIPCRNAMRGSFSLNGT

SEQ ID NO:50

>Cotton 11206330 GA\_\_EB0023J04F

130 CTCCGCCAGTGCAACTTGCTTAAAGTAGGGCCTAATGTTGGCAACAATGAACCTATCA  
TTGAGGAGCCTGCAACACCTGAACCCAGAGCATGCAGAAGGATCAGAGAGTGATATTGAAG  
ATGCAAGCTATGATGATCCAGATGAAATTTCCACAATAAACTCAACATTGAAGAGTTCA  
CAGCAAACCTACAGCATTACATGCAGGGCAATATGGAACCTCAAGAAGGGGACTTTGCA  
AGGCTTTAGTAGCTTTGAATCCTGATGCTGCTTCTATCCCTACTCCAAAATGAAGAATG

TAAGCAGGCTACGAACAGAGCACTATGTATATGAGCTTCCAGATAAACATCCTCTCTTGA  
AACAGATGGAAAAGCGGGAACCTGATGATCCTAGCCCCTATCTTCTTGCAATATGGACAC  
CAGGTGAAACTGCAAACTCAATTCAACCACCAGAACAAGTTGTGGGTCCCAAGAACAG  
GAAGACTGTGCAATGAGAAGACCTGCTTTGCTTGCAACAGTGTAAGAGAAGCTAACACTG  
5 AGACAGTCCGAGGAACCATCTTGATACCTTGTAGAAATGCAATGAGAGGAAGCTTTTCCC  
TTAATGGGACTTAATTTTCAAGTTAATGAGGTCTTTTGCAGATCATGAATCAAGCCTCAA  
CCCAATCGACGTTCCAAGGGGAATGGATTGGGAATTTAACAAGAACGAAGCTGTATACTTG  
GAACATCCTGGTTCATCAATATTTAAAGGACTTTTCGACGAGGGAA

10 SEQ ID NO:51

>Soybean 5606759

MGWVPLQPLPESLQLHLLLELYPVLESIQKYLWPRCLKLDQETLYELHYQMITFGKVFCTK  
SKPNCNACPMRAECRHFASAFASARFALPGPEQKSIVSTGNSVINQNPSEIISQLHLPP  
PENTAQEBEIQLTEVSRQLESKFEIYICQPIIEBPTPEPECLQESXTDIEDAVYEDSS

15

SEQ ID NO:52

>Soybean 5606759 SB95C12.

ACGAGCTTCCCGGTAGACACAAATGTCGGACGTATTGCCGTACGACTGGGATGGGTGCCT  
CTGCAGCCACTGCCTGAGTCACTGCAGTTGCATCTCCTAGAATTGTACCCGGTGTGGAG  
20 TCAATACAAAAATATCTCTGGCCTCGACTGTGCAAGCTAGATCAGGAACACTATATGAG  
CTACATTACCAGATGATTACATTGGAAAGGTCTTCTGTACAAAAAGCAAACCAATTGT  
AATGCATGCCCAATGAGAGCAGAATGTAGACACTTTGCTAGTGCAATTTGCAAGTGCAAGG  
TTTGCACTGCCTGGACCAGAGCAGAAGATATAGTTAGCACAACTGGAATAGTGTGATT  
AACCAGAACCCATCTGAAATCATCAGTCAGTTGCACTTGCCCTCCACCTGAGAACACAGCC  
25 CAAGAAGATGAAATTCAACTAACAGAAGTGAGCAGACAATTGGAATCAAAATTTGAAATA  
TATATTTGCCAACCTATCATTGAAGAGCCGAGAAGCTCCAGAGCCAGAATGCTTGCAAGAA  
TCACANACTGATATAGAGGATGCTGTCTATGAGGATTCAAGTG

SEQ ID NO:53

30 >Wheat 12019155

MFHCHGTRGSDLGFDLNTPEQKAPQRRKRPKVIKEAKPKSTRKPATQKTQMKENPHKK  
RKYVRKTAATPQTNVTEESVDSIVATKKSCRRALNFDLEHNKYASQSTISCCQEQEIDHRNE  
KAFNTTSDHKAKEPKNTDDNTLLHLEKQANNFQSE



SEQ ID NO:54

>Wheat 12019155

5 AACAGTCAGGACAAAGGCAACAAGATCAGCAGTCAGGACAAGGGCAGCAACCGGGACAAA  
GGCAGCCAGGGTACTACTCAACTTCTCCGCAACAATTAGGACAAGGCCAACCAAGGTACT  
ACCCAACTTCTCCGAGCAGCCAGGACAAGAGCAGCAGCCAAAGACAATTGCAACAACCAG  
AACAAGGGCAACAAGGTCAGCAGCCAGAAACAAGGGCAGCAAGGTGAGCAGCAAGACAAG  
GGGAGCAAGGTGAGCAGCCAGGACAAGGGCAACAAGGGCAGCAACCGGGACAAGGGCAGC  
CAGGGTACTACCCAACTTCTCCGAGCAGTCAGGACAAGGGCAACCAGGGTACTACCCAA  
10 CTTCTCCACAGCAGTCAGGACAATTGCAACAACCAGCACAAGGGCAGCAACCAGGACAAG  
AGCAACAAGGTCAACAGCCAGGACAAGGGCAGCAACCGGGACAAGGGCAAGCCAGGGTAC  
TACCCAACTTCTCCGAGCAGTCAGGACAAGAGCAACAGCTAGAACAATGGCAACAGTCA  
GGACAGGGGCAACCAGGGCACTACCCAACTTCTCCGTTGCAAGCCAGGACAAGGGCAACC  
AGGGTACTACCCAACTTCTCACAACAGATAGGACAAGGGCAGCAGCCAAAGACAATTTGC  
15 ACAACCAACACAAGGGCAACAANGGGCAGCAACCAAGGACAANGGGCAACAAGGTCAACA  
GCCANGAAAAAGGCAACAAGGTCAAGCAACCAAGNACAAGGGGCAAGCAANCCAGGAC  
AAGGGCAGCCANGTCTACCCAACTTNTTTGAGCAAGTCANGGAAAAGGGGCACCANC  
CNAGGANAAATGGGNACCACCCAGNACAAGGACAACCCCGGGTCTTCCCCAAANTTTTNC  
CN

20

SEQ ID NO:55

>Tomato 8106032

MSLAHFPLKTDSTQKHGNTGIIIEEPEECATDPNVSIWYEDQPNQSTHCQDSSGVYN  
TDSNEEKPAVNDSSESENSTECIKSAECSVILQSDSSREGSDLYHGSTVTSSQDRKELND  
25 LPSSPSSVVSSEISAVIQASEGTDSSNFCSSSTSLFLKLLQMGATSGAQGTRCTEHLHNQHK  
GNXGQQPRTXGNKVNSPXKKATKVKQPXTRGSXPQQGQPSYPTXFEQVXEKGHXPRXNG  
XHPXQQPVRVFPKXF

SEQ ID NO:56

30 >Tomato 8106032 EST356474

CTCGTGCCGGTTGGGGTATATCTTACACAGAATGTTTCAGATCACCTTTCTAGTTCTGCA  
TTCATGTCACTCGCTGCCCACTTTCTCTGAAAACAGACAGTACTCAGAAGCATGAAGGA  
AATACAGGTATTATAATTGAAGAACCCTGAAGAGTGTGCAACAGACCCCCAATGTTTCCATC  
AGATGGTATGAAGATCAACCAATCAGTCAACCCATTGTGAGGATCTTTCAGGAGTCTAT

AATACAGATTCAAATGAAGAAAAACCAGCTGTCAATGACTCTGAATCAAGTAAAAATAGC  
ACAGAATGCATAAAATCAGCAGAATGTTCTGTAAATCTGCAATCAGATTCTTCTAGAGAA  
GGCTCAGATCTGTATCATGGATCAACAGTTACAAGTTCCCAAGATCGAAAAGAGTTGAAT  
GATTGCGCTTCTTCTCCGAGTTCTGTTGTTTCTTCTGAGATCTCTGCTGTTATTCAAGCT  
5 TCAGAAGGAAGTACTCAAGCAACTTTTGCAGCTCCACTTCTTTTTTGAAGCTATTACAG  
ATGGCAGGAAGTTCAGGAGCACAAAGGAACCAGGTGCACTGAACATCTAC

SEQ ID NO:57

>Corn 1AW042334;

10 DAHPLLQQLGLDQREHDDPTPYLLAIWTPDGIKEITKTPKPCDPQMGDLNEMCHNC  
TAEKENQSRVYRGITLVP CRTAMRGSFPLNGTYFQVNEVFADHRSSHNP IHVEREMLWNL  
QRRMVFFGTSVPTIFKGLRTEEIQCFWRGFVCVRGFD METRAPRPLCPHLHVIARPKA

SEQ ID NO:58

15 >Corn 1AW042334 614027C01.y1 614 -

GAATTCGGCACCAGCAGATGCACATCCACTTTTACAACAGCTAGGACTTGACCAACGGGA  
ACATGATGATCCTACCCCACTTATTGGCCATATGGACACCAGATGGAATAAAGGAAAT  
AACTAAGACACCAAAACCATGCTGTGACCTCAAATGGAGGCGATTATGCAATAATGA  
AATGTGCCACAATTGTACTGCAGAGAAAGAAAACCAATCTAGATATGTCAGAGGCACAAT  
20 TCTGTTCCCTTGTCGAACAGCTATGAGGGGTAGTTTCCCACTTAATGGCACTTACTTTCA  
AGTCAATGAGGTATTGTCTGACCACAGATCTAGCCACAACCAATCCATGTGGAAAGGGA  
GATGCTATGGAACTTGCAAAGGCGCATGGTCTTTTTTCGGGACTTCAGTACCCACCATATT  
CAAAGGTCTAAGAACAGAAAGAAATACAACAATGCTTCTGGAGGGGATTGTCTGTGTGCG  
AGGATTGACATGGAGACTAGAGCACCAAGGCCTCTGTGCCCCCATTGACGCTTATAGC  
25 AAGGCCGAAAGCCCGCAAGACAGCAGCAACTGAGCAAGTACTCTAATCAGCAAAG

SEQ ID NO:59

>Corn AW076298

PCRTAMRGSFPLNGTYFQVNEVFADHCSSQNPIDVPRSWIWDLPRTVYFGTSVPTIFRG  
30 LSTEQIQFCFWKGFVCVRGFEQKTRAPRPLMARLHFPASKLKNKLTTEEIQCFWRGFV  
CVRGFDRTVRAPRPLYARLHFPASKVVRGK

SEQ ID NO:60

>Corn AW076298 614065C03.Y1 614 -

CGGCCCCAGACCATGCCGGACAGCAATGAGAGGAAGCTTCCCACTTAATGGGACATATTT  
TCAAGTTAATGAGGTATTTGCTGACCATTGTTCAAGCCAAAATCCAATTGATGTCCACG  
5 AAGTTGGATATGGGACCTCCCAAGACGAACGTTTTACTTTGGAACCTCAGTTCCTACAAT  
ATTTAGAGGTTTAACGACTGAAGAGATACAACAATGCTTTTGAGAGGATTGCTTTGTGT  
GAGGGGCTTTGATAGGACAGTAAGGGCACCAAGGCCCTTTATGCAAGGTTGCATTTTCC  
TGCCAGCAAGGTTGTTAGAGGCAAAAAGCCTGGAGCGGCAAGCGTCGAAGAATAATAGGT  
ACATCGAAGAAATATAGAGGAGCTAACAAAACGGATGGATAGCCCTAAATGAGATGCTGA  
10 CCCAATAAGTCGCCGAATCACCTCCAAGTTCTAACCCAAATTTTGGAGCGACATGACCTG  
TTAAATATATGTTCCATCTATGGTAACAGCTTAGATGTTCTTGAGTGCATATTTCTTA  
CTCTGAAATTCATATAGCAAATGAAAAAAACACAGTGCATAGTCTAGTTCTAATTGTA  
CCTGTGAGTGGAAATCAGTTGTTGTACAACATGAAGATGGG

15 SEQ ID NO:61

>Corn BE639158;

KNSEPIIEEPASPREERPETMENDIEDFYEDGEIPTIKLNMEAFQNLNENIKESNNEL  
QSDDIKALVAISTEAASIPVPKLKNVLRRLTEHYVYELPDHPLQLQLGLDQREHDDPT  
PYLLAIWTPDGIKEITKTPK

20

SEQ ID NO:62

>Corn BE639158 946021E09.Y1 946 -

TGAGCTGCATTATCAGATGATTACATTTGGAAAGGTCCTTTGTACCAAAAGACAGCCAAA  
TTGCAATGCATGCTATGAATTCGACTCACCTACCTCGCCTTGAGGGGAGTATCCATTCAA  
25 GGGAGTTTCTTCTCAAGAAATCAGAGCCAATAATCGAGGAGCCTGCAAGTCCAAGAGAG  
AAAGACCTCCAGAAACCATGGAAAATGATATTGAAGATTTTTATGAAGATGGTGAAATCC  
CAACAATAAAGCTTAACATGGAAGCTTTTGCACAAAAACCTGGAGAATTGCATTAAAGAAA  
GCAATAACGAACCTCCAGTCTGATGATATTGCAAAAAGCATTGGTGTCTATTAGCACTGAAG  
CAGCTTCGATTCTGTACCGAACTAAAGAATGTGCTTAGGCTTCGAACAGAACACATATG  
30 TGTATGAGCTTCCAGATGCACATCCACTTTTACAACAGCTAGGACTTGACCAACGGGAAC  
ATGATGATCTTACCCATACCTATTGGCCATATGGACACCAGATGGAATAAAGGAAATAA  
CTAAGACACCAAAACCATGCT

SEQ ID NO:63

>Corn T25243;

NHQPIIEEPLSPECETENIEAHEGAIEDFFCEESDEIPTINLNIEEFTQNLKDYMQANNV  
EIXYADMSKALVAITPDAASIPTPKLKNVNRRLRTEHQVYELPDSHPLEGFQXEPDDPC  
5 PYLLSIWTPGELHNRSMP

SEQ ID NO:64

>Corn T25243;

CTGGTAATCATCAGCCAATCATCGAGGAACCACTGAGCCCAGAATGTGAACTGAAAATA  
10 TAGAGGCACATGAGGGTGCAATTGAGGATTTCTTTTGTGAAGAATCTGATGAAATTCCTA  
CCATTAATCTTAATATCGAGGAGTTCACACAAAACCTGAAGGACTATATGCAAGCAAACA  
ATGTTGAGATTGANTATGCTGACATGTCAAAGGCATTGGTTGCCATCAGCCTGATGCTG  
CTTCCATTCCAACCTCAAAGCTCAAGAATGTCAATCGTCTGAGGACAGAACACCAAGTTT  
ATGAAGTCCAGATTACACCCCTCTTCTGGAAGGATTCGAACAGNGNGAACCAGATGATC  
15 CCTGTCCATATCTTCTTCCATATGGACCCAGGTGAAGTGCACAATCGATCGATGCCCC  
AA

SEQ ID NO:65

>Corn AW453174;

20 FQGNEVFADHCSRQNPIDGPRSWIWDLPRTGYFGTSGPTIFRGLTTEEIQRCFWRGFVC  
VRGFDRITVRAPRPLYARLHFPVSKVVRGK

SEQ ID NO:66

>Corn AW453174 660032D01.Y1 660 -;

25 CATGCCGAACAGCAATGAGAGGAAGCTTCCCACTTAATGGGACGATTTTCAAGGTAATGA  
GGTATTGTGCTGACCATTGCTCAAGGCAAAATCCAATTGATGGCCCACGAAGTTGGATTG  
GGACCTTCCAAGACGAACCTGTTTACTTTGGAACCTCAGGTCCTACATATTGAGAGGGTT  
AACGACTGAAGAGATACAACGATGCTTTTGGAGAGGATTTGTTTGCCTGAGGGGCTTTGA  
TAGGACAGTGCGGGCACCAAGGCCCTTTATGCAAGGTTGCATTTTCTGTCAGCAAGGT  
30 TGTTAGAGGCAAAAGCCTGGAGCAGCAAGAGCAGAAGAATAATAGAACATTGAAGAAAT  
ATAGGGGTGCTAACCAGATGAGGATGGATAGCCCGAAATGAGATGCTGACCCAATAGGTC  
GCCAAATCACCTCCAAATTCTAACCCAATGACTTCCATCTGTAATGAATGGCAATACCTT  
GAAAACCT

SEQ ID NO:67

>Corn BE509759;

NGTYFQVNEVFADHRSSHNP IHVEREMLWNLQRRMVFFGTSVPTIFKGLRTEEIQQCFWR  
GFVVCVRGFDMETRAPRPLCPHLII IARPKARKT

5

SEQ ID NO:68

>Corn BE509759 946021E09.X1 946 -

TGGCATCTTACATGGACTAACAGCTAGATGCTAATTTACATACAGTAGATCTGAAACAAAAAGTGAAAATTATTGGTGCT  
TTCCTGATGCTTCATTAGTCCTCTCGTCTCAGAACTAACAGTCTCGAACCCCATCCATGGCTTAAATTTTCCTAAACAAT  
10 GGCTCTTTTTAGGCAGGAAGTAATATGATTCATGCATAGGTCGAGAGCTATTGATGTCATATCACAATAAACATGATG  
TTCATAAACTGATATCTTTGCTGATTAGAGTACTTGCTCAGTTGCTGCTGCTTGGGGCCCTTCGGCCCTTGCTATAATG  
TGCAAAATGGGGCAGACAGAGGCCCTTGGTGCTCTAGTCTCCATGTCGAATCCTCGCACACAGACAAATCCCTCCAGAAGCA  
TTGTTGTAATTTCTTCTGTTCTTAGACCTTTGAATATGGTGGGTACTGAAGTCCCGAAAAAGACCATCGCGCTTTGCAAGT  
15 TCCATAGCATCTCCCTTCCACATGGATTGGGTTGTGCTAGATCTGTGCTCAGCAAAATACCTCATTGACTTGAAGTAA  
GTGCCATTAA

SEQ ID NO:69

>Corn 1AW017984;

VPRSWIWDLPRTTVYFGTSVPTIFRGLTTEEIQQCFWRGFVVCVRGFDRTVRAPRPLYARL  
20 HFPASKVVVRGK

SEQ ID NO:70

>Corn 1AW017984;

CCTGAAACAATCAAATAACGGCCGATGAGGTTACATTGTTTATAGTATATGATCAAAGAA  
25 CATGTATGACCAATTGTACAAATAGGCCCATCTTCATGTTGTACAACAAGTGAATCCACTC  
ACAGGTACAATTAGAAGTACTATGCACTGTGTTTTTTTCATTGTGCTATATGAATTT  
CAGAGTAAAGAAATATGCGACTCACAAGAACATCTAAGCTGTTACCATAGATGGAACATAA  
TTTAACAGGTCATGTGCGCTCAAAAAATGGGTTAGAAGTTGGAGGTGATTTCGGCGACTTA  
TTGGGTGAGCATCTCATTTAGGGCTATCCATCCGTTTTTGTAGCTCTCTATATTTCTTC  
30 GATGTACCTATTATTCTTCGACGCTTGCCGCTCCAGGCTTTTTGCGCTCTTAACAACCTTGC  
TGGCAGGAAAAATGCAACCTTGCCATAAAGGGGCCCTTGGTGCCCTTACTGTCTCTATCAAAGC  
CCCTCACACAAACGAATCCTCTCAAAGCATTTGTTGTATCTCTTCAGTCGTTAAACCTC  
TAAATATTGTAGGAAGTGAAGTTCCAAAGTAAACAGTTCGTCTTGGGAGGTCCCATATCC  
AACTTCGTGGGAC

35